

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

4(164)
2010

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ

Издательство "Новые технологии"

СОДЕРЖАНИЕ

БАЗЫ ДАННЫХ

- Кудрявцев К. Я. Спектральный метод поиска ключевых слов в полнотекстовых базах данных. 2
Полищук Ю. В., Черных Т. А. Построение хранилищ данных, ориентированных на работу с квазиструктурированным информационным наполнением документов. . . 9

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ И УПРАВЛЕНИЕ РИСКАМИ

- Дрюченко М. А., Сирота А. А. Математическое и программное обеспечение для решения задачи стегаанализа на основе статистических и нейросетевых алгоритмов обработки информации. 14
Кустов Г. А., Николаева М. А., Зотова О. Ф., Шарапов Р. А. Алгоритмы выбора превентивных мер снижения рисков. 22

НЕЙРОСЕТИ

- Жернаков С. В. Восстановление потерянной информации с датчиков ГТД в условиях бортовой реализации с использованием технологии нейронных сетей. . . 28
Глущенко О. И., Горбатков С. А. Нейросетевая модель поддержки принятия решений в процессах муниципального бюджетирования. 34

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Соболев С. С., Сотов Л. С., Харин В. Н. Алгоритм работы и модель функционального генератора перестановок. 41
Абдуллаев Н. Т., Алиев Т. А., Дышин О. А. Алгоритм вычисления статистической оценки обобщенных размерностей множеств. 46

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

- Львович Я. Е., Королев Е. Н. Преимущества объектно-ориентированного подхода для проектирования систем обучения в интеллектуальных системах принятия решений. 56
Пантелеев Е. Р., Карпов Я. Э. Разработка и интерпретация решения расчетных задач в среде Web-обучения. 59
Сафонов В. В. Метод ветвей и границ в прямой и обратной задачах оптимального распределения времени на изучение цикла дисциплин. 62

ДИСКУССИОННЫЙ КЛУБ

- Гуревич И. М. Информация как универсальная неоднородность. 66

ИНФОРМАЦИЯ

- Рожкова М. В. Мои заметки о Mobile World Congress 2010. Тезисно: значимые события и тренды. 75
VIII межрегиональная конференция "Информационные технологии и решения для «Электронной России» и IX окружная выставка "Информационные технологии XXI века". 77
Contents. 79
Приложение. Позин Б. А. Ввод в действие информационных систем и сопровождение их программного обеспечения

Главный редактор
НОРЕНКОВ И. П.

Зам. гл. редактора
ФИЛИМОНОВ Н. Б.

Редакционная
коллегия:

АВДОШИН С. М.
АНТОНОВ Б. И.
БАТИЩЕВ Д. И.
БАРСКИЙ А. Б.
БОЖКО А. Н.
ВАСЕНИН В. А.
ГАЛУШКИН А. И.
ГЛОРИОЗОВ Е. Л.
ГОРБАТОВ В. А.
ДОМРАЧЕВ В. Г.
ЗАГИДУЛЛИН Р. Ш.
ЗАРУБИН В. С.
ИВАННИКОВ А. Д.
ИСАЕНКО Р. О.
КОЛИН К. К.
КУЛАГИН В. П.
КУРЕЙЧИК В. М.
ЛЬВОВИЧ Я. Е.
МАЛЬЦЕВ П. П.
МЕДВЕДЕВ Н. В.
МИХАЙЛОВ Б. М.
НАРИНЬЯНИ А. С.
НЕЧАЕВ В. В.
ПАВЛОВ В. В.
ПУЗАНКОВ Д. В.
РЯБОВ Г. Г.
СОКОЛОВ Б. В.
СТЕМПКОВСКИЙ А. Л.
УСКОВ В. Л.
ЧЕРМОШЕНЦЕВ С. Ф.
ШИЛОВ В. В.

Редакция:

БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://www.informika.ru/text/magaz/it/> или <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

УДК 004.91

К. Я. Кудрявцев, канд. техн. наук, доц.,
Московский инженерно-физический
институт (ГУ),
e-mail: const58@mail.ru

Спектральный метод поиска ключевых слов в полнотекстовых базах данных

Предлагается оригинальный подход поиска ключевых слов в полнотекстовых базах данных, основанный на теории спектрального анализа и использовании алгоритма быстрого преобразования Фурье.

Ключевые слова: полнотекстовые базы данных, быстрое преобразование Фурье, поиск по ключевым словам в полнотекстовых базах данных

Поиск требуемой информации в полнотекстовых базах данных осуществляется по так называемым ключевым словам, которые определяются библиотекарем или автором и отражают основные положения печатной работы. Задание ключевых слов определяется на этапе внесения информации о печатном издании в базу данных и является довольно ответственным моментом, так как отсутствие какого-либо ключевого слова приведет к тому, что информация о печатной работе не будет выведена поисковыми системами. Аналогичные подходы используются поисковыми системами сети Интернет, когда информация о сайте включается в ответ на запрос при наличии определенных ключевых слов, связанных с сайтом и заданных разработчиком сайта.

Как видно, проблема заключается в необходимости предварительного задания набора ключевых слов, и отсутствие некоторых из них исключает печатную работу из списка поиска. Было бы интересно создать такую систему поиска информации в полнотекстовых базах данных, для которой не требовалось бы предварительного задания списка ключевых слов.

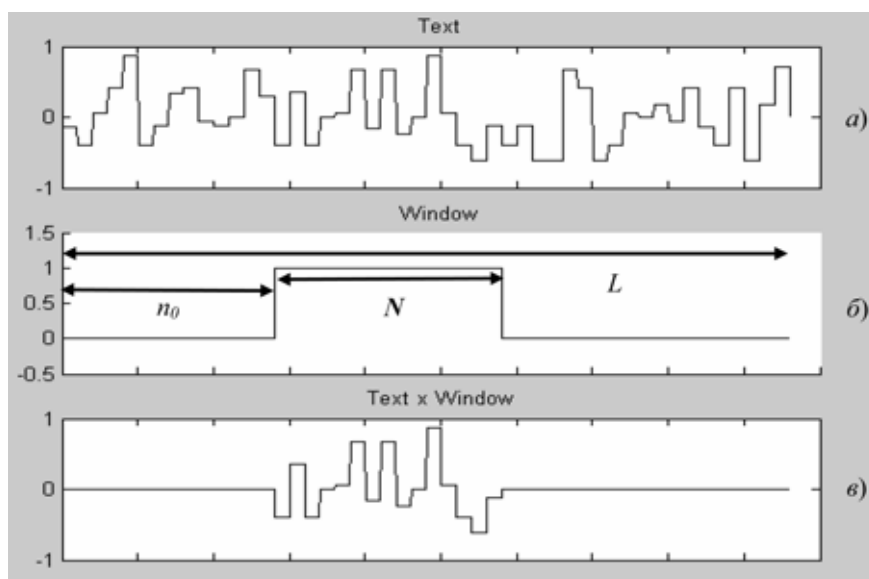
В данной работе предлагается новый подход создания поисковых систем без предварительного задания набора ключевых слов для каждой печатной работы. Подход строится на основе теории спектрального анализа (теории дискретного преобразования Фурье) [1] с использованием методов быстрого преобразования Фурье (БПФ).

Вывод основных соотношений с использованием "оконной" функции

Представим текст из файла в виде ряда $\{x(n)\}$ длиной L ($n = 0, 1, \dots, L-1$). Текст целесообразно привести к нижнему регистру и удалить из него пробелы, знаки препинания и прочие незначащие символы [2]. Значениями $x(n)$ можно считать, например, ASCII-коды символов (хотя можно использовать и собственную кодировку, например, числа из диапазона $[-1, +1]$). На представленном рис. 1, а изображен именно такой текстовый ряд $\{x(n)\}$. Будем считать, что ключевое слово, имеющее длину N , присутствует в тексте и расположено, начиная с позиции n_0 .

Построим ряд для "оконной" функции $\{w(n)\}$ длины L , имеющий нулевые значения всюду, кроме интервала $[n_0, n_0 + 1, \dots, n_0 + N - 1]$. Данный ряд изображен на рис. 1, б:

$$w(n) = \begin{cases} 1, & n \in [n_0, n_0 + N - 1]; \\ 0, & n \notin [n_0, n_0 + N - 1]. \end{cases} \quad (1)$$



Представление текста в виде сигнала $\{x(n)\}$

Произведение рядов $\{x(n) \cdot w(n)\}$ (результат произведения изображен на рис. 1, в) позволяет выделить ключевое слово.

Используя дискретное преобразование Фурье (ДПФ), вычислим спектр ряда $\{x(n) \cdot w(n)\}$ ($n = 0, 1, \dots, L - 1$), который можно представить как свертку спектров $X(p)$, $W(p)$:

$$\sum_{n=0}^{L-1} x(n)w(n)l^{-j\frac{2\pi}{L}kn} = \frac{1}{L} \sum_{p=0}^L X(p)W(k-p), \quad (2)$$

где $X(p)$ — спектр исходного текста, а $W(p)$ — спектр "оконной" функции.

Спектр "оконной" функции довольно легко вычисляется:

$$W(k) = \sum_{n=0}^{L-1} w(n)l^{-j\frac{2\pi}{L}kn} = \sum_{n=n_0}^{n_0+N-1} l^{-j\frac{2\pi}{L}kn}. \quad (3)$$

Выполняя замену переменной $n = p + n_0$, запишем (3) в виде

$$\begin{aligned} \sum_{n=n_0}^{n_0+N-1} l^{-j\frac{2\pi}{L}kn} &= \sum_{p=0}^{N-1} l^{-j\frac{2\pi}{L}k(p+n_0)} = \\ &= l^{-j\frac{2\pi}{L}kn_0} \sum_{p=0}^{N-1} l^{-j\frac{2\pi}{L}kp}. \end{aligned} \quad (4)$$

Выражение в правой части (4) представляет собой сумму геометрической прогрессии, которую

можно найти, учитывая, что $a_0 = 1$, $q = l^{-j\frac{2\pi}{L}k}$:

$$\begin{aligned} W(k) &= l^{-j\frac{2\pi}{L}kn_0} \frac{1 - l^{-j\frac{2\pi}{L}kN}}{1 - l^{-j\frac{2\pi}{L}k}} = \\ &= l^{-j\frac{2\pi}{L}kn_0} \frac{l^{-j\frac{\pi}{L}kN} \left(l^{j\frac{\pi}{L}kN} - l^{-j\frac{\pi}{L}kN} \right)}{l^{-j\frac{\pi}{L}k} \left(l^{j\frac{\pi}{L}k} - l^{-j\frac{\pi}{L}k} \right)} = \\ &= l^{-j\frac{2\pi}{L}kn_0} l^{-j\frac{2\pi}{L}k(N-1)} \frac{\sin \frac{\pi}{L}Nk}{\sin \frac{\pi}{L}k}. \end{aligned}$$

Таким образом, получим

$$W(k) = l^{-j\frac{2\pi}{L}kn_0} l^{-j\frac{\pi}{L}k(N-1)} \frac{\sin \frac{\pi}{L}Nk}{\sin \frac{\pi}{L}k}. \quad (5)$$

В свою очередь, левая часть выражения (2) может быть записана как

$$\begin{aligned} \sum_{n=0}^{L-1} x(n)w(n)l^{-j\frac{2\pi}{L}kn} &= \sum_{n=n_0}^{n_0+N-1} x(n)l^{-j\frac{2\pi}{L}kn} = \\ &= \sum_{p=0}^{N-1} x(n_0+p)l^{-j\frac{2\pi}{L}k(n_0+p)} = \\ &= l^{-j\frac{2\pi}{L}kn_0} \sum_{p=0}^{N-1} x(n_0+p)l^{-j\frac{2\pi}{L}kp}. \end{aligned} \quad (6)$$

Подставляя (5) и (6) в (2) и сокращая правую и левую часть на $l^{-j\frac{2\pi}{L}kn_0}$, получим базовое соотношение

$$\begin{aligned} \sum_{p=0}^{N-1} x(n_0+p)l^{-j\frac{2\pi}{L}kp} &= \\ &= \frac{1}{L} \sum_{p=0}^L X(p)l^{-j\frac{\pi}{L}(N-1)(k-p)} \frac{\sin \frac{\pi}{L}N(k-p)}{\sin \frac{\pi}{L}(k-p)} l^{j\frac{2\pi}{L}pn_0}. \end{aligned} \quad (7)$$

Полагая в (7) $k = 0$, будем иметь

$$\begin{aligned} \sum_{p=0}^{N-1} x(n_0+p) &= \\ &= \frac{1}{L} \sum_{p=0}^L X(p)l^{j\frac{\pi}{L}(N-1)p} \frac{\sin \frac{\pi}{L}Np}{\sin \frac{\pi}{L}p} l^{j\frac{2\pi}{L}pn_0}. \end{aligned} \quad (8)$$

Введя обозначение

$$Y_0(p) = X(p)l^{j\frac{\pi}{L}(N-1)p} \frac{\sin \frac{\pi}{L}Np}{\sin \frac{\pi}{L}p}, \quad (9)$$

запишем (8) в виде

$$\sum_{p=0}^{N-1} x(n_0+p) = \frac{1}{L} \sum_{p=0}^L Y_0(p)l^{j\frac{2\pi}{L}pn_0} = y_0(n_0). \quad (10)$$

Правая часть соотношения (10) представляет собой выражение для обратного дискретного преобразования Фурье и может быть вычислено с помощью алгоритма инверсного БПФ.

Выполняя аналогичные манипуляции для $k = L/2$ (L должно быть четным; если L нечетное, то дополним текст маловероятным и редко встречающимся символом, например "Ь"), получим

$$\begin{aligned} \sum_{p=0}^{N-1} (-1)^n x(n_0 + p) &= \\ &= \frac{l^{-j\frac{\pi}{2}(N-1)}}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N-1)p} \times \\ &\times \frac{\sin\left(\frac{\pi}{2}N - \frac{\pi}{L}Np\right)}{\cos\frac{\pi}{L}p} l^{j\frac{2\pi}{L}pn_0} = y_{\frac{L}{2}}(n_0). \end{aligned} \quad (11)$$

Введя обозначение

$$Y_{\frac{L}{2}}(p) = X(p) l^{j\frac{\pi}{L}(N-1)p} \frac{\sin\left(\frac{\pi}{2}N - \frac{\pi}{L}Np\right)}{\cos\frac{\pi}{L}p}, \quad (12)$$

представим (11) в виде

$$\begin{aligned} \sum_{p=0}^{N-1} (-1)^n x(n_0 + p) &= \\ &= \frac{l^{-j\frac{\pi}{2}(N-1)}}{L} \sum_{p=0}^L Y_{\frac{L}{2}}(p) l^{j\frac{2\pi}{L}pn_0} = y_{\frac{L}{2}}(n_0), \end{aligned} \quad (13)$$

которое также представляет собой выражение для обратного дискретного преобразования Фурье и может быть вычислено с помощью алгоритма инверсного БПФ.

Можно показать, что при вычислении правых частей выражений (10) и (13) остаются только действительные значения комплексных чисел, а мнимые составляющие равны нулю. Рассмотрим сначала формулу (10) и запишем правую часть как

$$\begin{aligned} \frac{1}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N-1)p} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p} l^{j\frac{2\pi}{L}pn_0} &= \\ &= \frac{1}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p}. \end{aligned} \quad (14)$$

Рассмотрим пару слагаемых ($p = 1, 2, \dots, L/2$)

$$X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p} \quad (15)$$

и

$$\begin{aligned} X(L-p) l^{j\frac{\pi}{L}(N+2n_0-1)(L-p)} \frac{\sin\frac{\pi}{L}N(L-p)}{\sin\frac{\pi}{L}(L-p)} &= \\ &= X(L-p) l^{j2\pi n_0} l^{j\pi(N-1)} \times \\ &\times l^{-j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}N(L-p)}{\sin\frac{\pi}{L}(L-p)}. \end{aligned} \quad (16)$$

Представляя $\frac{\sin\frac{\pi}{L}N(L-p)}{\sin\frac{\pi}{L}(L-p)}$ как

$$\begin{aligned} \frac{\sin\frac{\pi}{L}N(L-p)}{\sin\frac{\pi}{L}(L-p)} &= \frac{\sin\left(\pi N - \frac{\pi}{L}Np\right)}{\sin\left(\pi - \frac{\pi}{L}p\right)} = \\ &= \frac{\cos\pi N \sin\frac{\pi}{L}Np}{\cos\pi \sin\frac{\pi}{L}p} = (-1)^{N-1} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p}, \end{aligned}$$

а также учитывая, что $l^{j2\pi n_0} = 1$, $l^{j\pi(N-1)} = (-1)^{N-1}$ и $(-1)^{2(N-1)} = 1$, перепишем последнее выражение в виде

$$\begin{aligned} X(L-p) l^{j2\pi n_0} l^{j\pi(N-1)} \times \\ \times l^{-j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}N(L-p)}{\sin\frac{\pi}{L}(L-p)} &= \\ &= X(L-p) l^{-j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p}. \end{aligned} \quad (17)$$

Так как $X(p)$ — спектр действительного сигнала, то модуль $|X(p)| = |X(L-p)|$, а фаза $\varphi(p) = -\varphi(L-p)$. Полагая, что L является четным (если L — нечетное, то дополним текст маловероятным и редко встречающимся символом, например "ъ"), представим (14) в виде

$$\begin{aligned} \frac{1}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin\frac{\pi}{L}Np}{\sin\frac{\pi}{L}p} &= \\ &= \frac{1}{L} \left\{ NX(0) + X\left(\frac{L}{2}\right) l^{j\frac{\pi}{2}(N+2n_0-1)} \sin\frac{\pi}{2}N + \right. \end{aligned}$$

$$\begin{aligned}
& + \sum_{p=1}^{\frac{L}{2}-1} X(p) l^{j\frac{\pi}{2}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} + \\
& + \sum_{p=1}^{\frac{L}{2}-1} X(L-p) l^{-j\frac{\pi}{2}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} \Bigg\} = \\
& = \frac{1}{L} \left\{ NX(0) + X\left(\frac{L}{2}\right) l^{j\frac{\pi}{2}(N+2n_0-1)} \sin \frac{\pi}{2} N + \right. \\
& + \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} + \\
& + \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{-j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} \Bigg\} = \\
& = \frac{1}{L} \left\{ NX(0) + X\left(\frac{L}{2}\right) l^{j\frac{\pi}{2}(N+2n_0-1)} \sin \frac{\pi}{2} N + \right. \\
& + \sum_{p=1}^{\frac{L}{2}-1} |X(p)| \left(l^{j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} + \right. \\
& + \left. l^{-j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} \right) \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} \Bigg\} = \\
& = \frac{1}{L} \left\{ NX(0) + X\left(\frac{L}{2}\right) l^{j\frac{\pi}{2}(N+2n_0-1)} \sin \frac{\pi}{2} N + \right. \\
& + 2 \sum_{p=1}^{\frac{L}{2}-1} |X(p)| \cos \left(\frac{\pi}{L} (N+2n_0-1)p + \right. \\
& \quad \left. + \varphi(p) \right) \frac{\sin \frac{\pi}{L} Np}{\sin \frac{\pi}{L} p} \Bigg\}.
\end{aligned}$$

Первое и третье слагаемые в данном выражении представляют собой действительные значения; $X(L/2)$ также является действительным. Поэтому остается выяснить, является ли

$l^{j\frac{\pi}{2}(N+2n_0-1)} \sin \frac{\pi}{2} N$ действительным значением.

В этом нетрудно убедиться, находя его мнимую часть, которая равна (учитывая, что $l^{j\varphi} = \cos \varphi + j \sin \varphi$):

$$\begin{aligned}
j \sin \frac{\pi}{2} (N+2n_0-1) \sin \frac{\pi}{2} N &= j \cos \pi n_0 \sin \frac{\pi}{2} (N-1) \times \\
&\times \sin \frac{\pi}{2} N = (-1)^{n_0} \left[\cos \frac{\pi}{2} - \cos \frac{\pi}{2} (2N-1) \right] = 0.
\end{aligned}$$

Таким образом, при вычислении по формуле (10) получается действительное значение.

Перейдем к рассмотрению формулы (13) и запишем ее в виде

$$\begin{aligned}
& \frac{l^{-j\frac{\pi}{2}(N-1)}}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \times \\
& \times \frac{\sin \left(\frac{\pi}{2} N - \frac{\pi}{L} Np \right)}{\sin \frac{\pi}{L} p}. \quad (18)
\end{aligned}$$

При четном и нечетном N возможны разные варианты, поэтому анализируем их по отдельности.

N — **четное**, тогда

$$\begin{aligned}
l^{-j\frac{\pi}{2}(N-1)} &= (-j)^{N-1} = (-1)^{N-1} j^{N-1} = \\
&= -j^{N-1} = jj^N = j(j^2)^{\frac{N}{2}} = j(-1)^{\frac{N}{2}} = (-1)^{\frac{N}{2}} j; \\
\sin \left(\frac{\pi}{2} N - \frac{\pi}{L} Np \right) &= -\cos \frac{\pi}{2} N \sin \frac{\pi}{L} Np = \\
&= -(-1)^{\frac{N}{2}} \sin \frac{\pi}{L} Np = (-1)^{\frac{N}{2}+1} \sin \frac{\pi}{L} Np.
\end{aligned}$$

Подставляя полученные выражения в (18), получаем

$$\begin{aligned}
& \frac{(-1)^{\frac{N}{2}} j (-1)^{\frac{N}{2}+1}}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} = \\
& = -\frac{j}{L} \sum_{p=0}^L X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p}. \quad (19)
\end{aligned}$$

Как и при анализе соотношения (10), рассмотрим пару слагаемых

$$X(p) l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p}$$

$$\begin{aligned}
& X(L-p) l^{j \frac{\pi}{L}(N+2n_0-1)(L-p)} \frac{\sin \frac{\pi}{L} N(L-p)}{\cos \frac{\pi}{L}(L-p)} = \\
& = X(L-p) l^{j \pi(N-1)} l^{j 2 \pi n_0} l^{-j \frac{\pi}{L}(N+2n_0-1)p} \times \\
& \quad \times \frac{\sin \frac{\pi}{L} N(L-p)}{\cos \frac{\pi}{L}(L-p)} = (-1)^{N-1} X(L-p) \times \\
& \quad \times l^{-j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \left[\pi N - \frac{\pi}{L} Np \right]}{\cos \left[\pi - \frac{\pi}{L} p \right]} = \\
& = -X(L-p) l^{-j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p}.
\end{aligned}$$

Используя полученные выражения и учитывая, что $|X(p)| = |X(L-p)|$, $\varphi(p) = -\varphi(L-p)$, представим (18) в виде

$$\begin{aligned}
& -\frac{j}{L} \sum_{p=0}^L X(p) l^{j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} = \\
& = -\frac{j}{L} \left\{ j \cdot (-1)^{\frac{N}{2}+n_0+1} NX\left(\frac{L}{2}\right) + \right. \\
& \quad + \sum_{p=1}^{\frac{L}{2}-1} X(p) l^{j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} - \\
& \quad \left. - \sum_{p=1}^{\frac{L}{2}-1} X(L-p) l^{-j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} \right\} = \\
& = -\frac{j}{L} \left\{ j \cdot (-1)^{\frac{N}{2}+n_0+1} NX\left(\frac{L}{2}\right) + \right. \\
& \quad + \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{j \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right]} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} - \\
& \quad \left. - \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{-j \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right]} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} \right\} =
\end{aligned}$$

$$\begin{aligned}
& - \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{-j \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right]} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} \left\} = \right. \\
& = \frac{j}{L} \left\{ j \cdot (-1)^{\frac{N}{2}+n_0+1} NX\left(\frac{L}{2}\right) + \right. \\
& \quad + \sum_{p=1}^{\frac{L}{2}-1} |X(p)| \left[l^{j \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right]} - \right. \\
& \quad \left. - l^{-j \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right]} \right] \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} \left\}.
\end{aligned}$$

Учитывая, что $\sin \varphi = \frac{l^j \varphi - l^{-j} \varphi}{2j}$, получим

$$\begin{aligned}
& -\frac{j}{L} \sum_{p=0}^L X(p) l^{j \frac{\pi}{L}(N+2n_0-1)p} \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} = \\
& = \frac{1}{L} \left\{ (-1)^{\frac{N}{2}+n_0} NX\left(\frac{L}{2}\right) + 2 \sum_{p=1}^{\frac{L}{2}-1} |X(p)| \times \right. \\
& \quad \times \sin \left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p) \right] \frac{\sin \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p} \left\}.
\end{aligned}$$

Как видно, данное выражение является действительным.

N — нечетное, тогда

$$\begin{aligned}
l^{-j \frac{\pi}{2}(N-1)} & = (-j)^{N-1} = (-1)^{N-1} j^{N-1} = j^{N-1} = \\
& = (j^2)^{\frac{N-1}{2}} = (-1)^{\frac{N-1}{2}}; \\
\sin \left(\frac{\pi}{2} N - \frac{\pi}{L} Np \right) & = \sin \frac{\pi}{2} N \cos \frac{\pi}{L} Np = \\
& = (-1)^{\frac{N-1}{2}} \cos \frac{\pi}{L} Np
\end{aligned}$$

и выражение (18) запишется как

$$\frac{1}{L} \sum_{p=0}^L X(p) l^{j \frac{\pi}{L}(N+2n_0-1)p} \frac{\cos \frac{\pi}{L} Np}{\cos \frac{\pi}{L} p}.$$

Пара слагаемых в данном случае будет выглядеть как

$$X(p)l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p};$$

$$X(L-p)l^{j\frac{\pi}{L}(N+2n_0-1)(L-p)} \frac{\cos\frac{\pi}{L}N(L-p)}{\cos\frac{\pi}{L}(L-p)} =$$

$$= X(L-p)l^{j\pi(N+2n_0-1)} l^{-j\frac{\pi}{L}(N+2n_0-1)p} \times \\ \times \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} = X(L-p)l^{-j\frac{\pi}{L}(N+2n_0-1)p} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p}.$$

Учитывая, что $|X(p)| = |X(L-p)|$, $\varphi(p) = -\varphi(L-p)$, получаем:

$$\frac{1}{L} \sum_{p=0}^L X(p)l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} = \\ = \frac{1}{L} \left\{ X(0) + X\left(\frac{L}{2}\right)N + \sum_{p=1}^{\frac{L}{2}-1} X(p) \times \right. \\ \times l^{j\frac{\pi}{L}(N+2n_0-1)p} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} + \\ \left. + \sum_{p=1}^{\frac{L}{2}-1} X(L-p)l^{-j\frac{\pi}{L}(N+2n_0-1)p} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} \right\} =$$

$$= \frac{1}{L} \left\{ X(0) + X\left(\frac{L}{2}\right)N + \right.$$

$$+ \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} +$$

$$+ \sum_{p=1}^{\frac{L}{2}-1} |X(p)| l^{-j\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right]} \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} \left. \right\} =$$

$$= \frac{1}{L} \left\{ X(0) + X\left(\frac{L}{2}\right)N + 2 \sum_{p=1}^{\frac{L}{2}-1} |X(p)| \times \right. \\ \times \cos\left[\frac{\pi}{L}(N+2n_0-1)p + \varphi(p)\right] \frac{\cos\frac{\pi}{L}Np}{\cos\frac{\pi}{L}p} \left. \right\}.$$

Как видно, представленное выражение является действительным, что и требовалось доказать.

Корреляционный подход к поиску ключевого слова

Найдем взаимную корреляционную функцию между рядом $\{x(n)\}$ длиной L и ключом $\{k(n-n_0)\}$ длиной N ($n = 0, 1, \dots, L-1$), где n_0 — смещение ключа относительно начала текста:

$$r(n_0) = \sum_{n=0}^{L-1} x(n)k(n-n_0), \quad n_0 = 0, 1, \dots, L-1.$$

Найдем ДПФ ряда $r(n_0)$. В соответствии с определением

$$R(m) = \sum_{n_0=0}^{L-1} r(n_0)l^{-j\frac{2\pi}{L}mn_0} = \\ = \sum_{n_0=0}^{L-1} \sum_{n=0}^{L-1} x(n)k(n-n_0)l^{-j\frac{2\pi}{L}mn_0} = \\ = \sum_{n_0=0}^{L-1} \sum_{n=0}^{L-1} x(n)l^{-j\frac{2\pi}{L}mn} k(n-n_0)l^{j\frac{2\pi}{L}m(n-n_0)} = \\ = X(m)K^*(m),$$

где $K^*(m)$ — комплексно-сопряженное к $K(m)$.

Таким образом,

$$R(m) = X(m)K^*(m). \quad (20)$$

Если в тексте имеется ключевое слово, начиная с позиции n_0 , то

$$r(n_0) = \sum_{n=0}^{N-1} k^2(n), \quad n_0 = 0, 1, \dots, L-N-1.$$

Следовательно, если среди значений $\{r(n_0)\}$

имеется значение, равное $\sum_{n=0}^{N-1} k^2(n)$, то ключевое

слово присутствует в тексте. Значения $\{r(n_0)\}$ для заданного ключевого слова легко могут быть найдены с помощью обратного быстрого преобразования Фурье спектрального ряда $X(m)K^*(m)$ $m = 0, 1, \dots, L-1$. Данное соотношение может выступать третьим дополнительным условием поиска ключевых слов в тексте.

Следует отметить, что значение $r(n_0)$ при ключевом слове не обязательно будет иметь максимальное значение в ряде $\{r(n_0)\}$.

Спектральный алгоритм поиска ключевых слов

На основании полученных соотношений (8)—(13) и (20), можно предложить следующий алгоритм поиска ключевых слов в файлах или в полнотекстовых базах данных.

1. Создание спектрограммы $\{X(p)\}$ ($p = 0, 1, \dots, L - 1$, где L — длина текста) исходного, модифицированного текста. Данную операцию следует проделать один раз на начальном этапе построения базы данных или при создании файла.

2. Вычисление выражений для заданного ключевого слова длины N

$$\begin{aligned} s &= \sum_{n=0}^{N-1} k(n), \quad p = \sum_{n=0}^{N-1} (-1)^n k(n), \\ q &= \sum_{n=0}^{N-1} k^2(n), \end{aligned} \quad (21)$$

где $\{k(n)\}$ — коды символов ключа длины N .

3. Вычисление по формулам (9), (12) и (20) "дополнительных" спектрограмм $\{Y_0(m)\}$, $\{Y_{L/2}(m)\}$ и $\{R(m)\}$ ($m = 0, 1, \dots, L - 1$, где L — длина текста).

4. Нахождение с помощью инверсного БПФ $\{y_0(n_0)\}$, $\{y_{L/2}(n_0)\}$ и $\{r(n_0)\}$,

$$n_0 = 0, 1, \dots, L - N - 1.$$

5. Проверка условий:

$$\begin{cases} s = y_0(n_0); \\ p = y_{L/2}(n_0); \\ q = r(n_0), \end{cases} \quad n_0 = 0, 1, \dots, L - N - 1. \quad (22)$$

Если при каком-либо n_0 условия (22) выполнены, то ключевое слово присутствует в тексте.

Кроме того, значение n_0 , при котором выполняются условия (22), определяет и место положения ключевого слова в тексте, что в ряде случаев является важным.

Заключение

Анализ современных поисковых систем, работающих с текстом, показывает, что они строятся

на индексировании текста, т. е. построении специального словаря, состоящего из всех слов текстового файла. Сама по себе процедура построения такого словаря является очень громоздкой и требует большого количества ресурсов. Помимо этого требуется применять сложные процедуры лексического и морфологического анализа, так как одно и то же слово может иметь много окончаний, словоформ и т. д.

Предлагаемый метод работает со спектрограммой текста, в которой в новом качественном виде содержится текстовый файл. Само понятие словоформ отходит на второй план и для поиска ключевых слов выполняются операции сложения и умножения.

Для уменьшения времени поиска возможно создание для каждого текстового документа двух матриц $\{y_0(n_0, N)\}$ и $\{y_{L/2}(n_0, N)\}$, $n_0 = 0, 1, \dots, L - N - 1$; $N = 3, 4, \dots, 15$, вычисленных для различных значений длины ключа. Для заданного ключа длины N подсчитываются значения по формуле (20), выбираются соответствующие столбцы матриц $\{y_0(n_0, N)\}$ и $\{y_{L/2}(n_0, N)\}$ и проводятся сравнения по формуле (20) с элементами выбранных столбцов. Если условие (21) выполняется, то ключевое слово имеется в тексте и данный текст включается в результирующий список поисковой системы.

В ряде случаев могут быть ошибки. Например, слова "МИР" и "РИМ" будут иметь одинаковые значения s , p и q , вычисленные по выражениям (21), но при реальном поиске обычно задаются довольно длинные слова (словосочетания) и вероятность совпадений становится небольшой. При этом следует отметить, что лучше выдать избыточную информацию, чем пропустить какой-либо документ.

Все полученные соотношения прошли экспериментальную проверку в системе MatLab 6.5, которая подтвердила их полную корректность.

Список литературы

1. Сергиенко А. Б. Цифровая обработка сигналов. СПб.: Питер, 2006. 751 с.
2. Кудрявцев Д. К., Кудрявцев К. Я. Спектральный метод поиска данных в текстовых файлах. Научная сессия МИФИ-2008 // Сб. научных трудов в 15 томах. Т. 11. Технологии разработки программных систем. Информационные технологии. М.: МИФИ. 2008. С. 190—191.

Ю. В. Полишук, канд. тех. наук, ст. препод.,
e-mail: youra_polishuk@bk.ru,
Т. А. Черных, аспирант,
Оренбургский государственный университет

Построение хранилищ данных, ориентированных на работу с квазиструктурированным информационным наполнением документов*

Рассмотрен подход к созданию хранилищ данных, ориентированных на работу с электронными документами, информационное наполнение которых квазиструктурировано. Предложена математическая интерпретация подхода.

Ключевые слова: хранение информации, квазиструктурированная информация, автоматизированные информационные системы

Введение

В зависимости от характера обрабатываемых информационных ресурсов автоматизированные информационные системы подразделяются на два основных класса: документальные и фактографические [1].

Документальные системы используются для работы с документами на естественном языке. Системы этого класса используются, как правило, для накопления и поиска документов по различным критериям.

Фактографические системы оперируют фактическими сведениями, представленными в виде специальным образом организованных совокупностей формализованных записей данных. Эти системы используются не только для реализации справочных функций, но и для решения задач обработки данных.

Перспективной является разработка автоматизированной системы, совмещающей в себе преимущества документальных и фактографических систем.

Такая необходимость обусловлена отсутствием жесткой связи между документом и извлеченными из него данными. При обработке документа извлеченные из него данные сохраняются в специально разработанной модели данных, и связь с

документом-первоисточником ослабевает или пропадает совсем.

Стоит отметить, что информация, извлекаемая из документа, зачастую представляет собой квазиструктурированные данные. Под квазиструктурированными данными обычно понимают информацию, в которой можно выделить некую структуру, однако структура эта заранее целиком или частично неизвестна либо может меняться с течением времени [2, 3].

Постановка задачи

Предлагаемая методика имеет следующую математическую интерпретацию.

Основой системы является реляционная база данных, схему которой можно описать следующим образом:

$$REL(R) = \{rel(r) | rel(r) = \{A_1, A_2, \dots, A_n\}\},$$

где $rel(r) = (A_1, A_2, \dots, A_n)$ — схема отношений, которая формируется перечнем и последовательностью атрибутов кортежа; A_i — атрибуты кортежей.

Для выбора требуемой записи из базы данных используется оператор выбора $\delta_B(r)$, который позволяет извлекать из отношения r кортежи, удовлетворяющие условию B , т. е.

$$r' = \delta_B(r) = \{t' | t' \subseteq r; B, rel(r') = rel(r)\} \subseteq r.$$

При этой операции сохраняется первоначальная схема отношений.

В случае, когда в качестве одного из атрибутов кортежа A_i выступает XML-документ, описываемый совокупностью квазиструктурированных моделей данных, в качестве параметра в условии B может быть использовано информационное наполнение XML-документа, получение которого осуществляет функция $xpath$ следующего вида:

$$xpath(path, S_{ns}) = sObj,$$

здесь S_{ns} — квазиструктурированная модель данных, относящаяся к пространству имен $namespace$, которую можно представить как совокупность множеств:

$$S_{ns} = \langle root, sObj, fObj, sMet, Obj_met \rangle,$$

где

$root$ — корневой объект, $root \in sObj$;

$sObj$ — конечное множество объектов;

$fObj$ — множество расширенных функциональных зависимостей на модели данных S_{ns} ;

$sMet$ — множество, такое что $sMet = rMet \times \times IMet$, $rMet$ — множество метасвойств оформления, $IMet$ — множество метасвойств ограничений;

* Работа выполнена в рамках исследовательской работы по ФЦП НПК-136 "Разработка технологии хранения и обработки квазиструктурированных данных".

Obj_met — функция, определенная на множестве $sObj$, такая что $sObj \xrightarrow{Obj_met} sMet$; каждому объекту $obj \in sObj$ ставится в соответствие кортеж из множества $sMet$, компоненты которого представляют собой метасвойства;

$path = fObj(sObj)$ — путь до искомого объекта квазиструктурированной модели S_{ns} ;

$namespace$ — пространство имен, т. е. некоторое множество, под которым подразумевается модель или окружение, созданное для логической группировки уникальных идентификаторов. Идентификатор, определенный в пространстве имен, ассоциируется с этим пространством. Один и тот же идентификатор может быть независимо определен в нескольких пространствах.

XML-документ Doc представлен как совокупность квазиструктурированных моделей данных:

$$Doc = \bigcup_{i=1}^n S_{ns_i},$$

где $i \in \{1, 2, \dots, n\}$ и S_{ns_i} — экземпляр квазиструктурированной модели данных.

Функция $xpath$ может быть использована не только в качестве условия выбора, но и в качестве нового атрибута в схеме отношений, получаемой в результате использования оператора выбора $\delta_B(r)$.

Практическая реализация

Описанная модель данных может быть реализована на технологии XML [4]. Рассмотрим подробнее ее возможности.

XML (eXtensible Markup Language — расширяемый язык разметки) фактически представляет собой свод общих синтаксических правил. XML — текстовый формат, предназначенный для хранения структурированных данных и обмена информацией между программами.

Целью создания XML было обеспечение совместимости при передаче структурированных данных между разными системами обработки информации, в частности, при передаче таких данных через сеть Интернет.

При работе с XML-документами стандартом определены два уровня "правильности":

- правильно построенный (Well-formed);
- действительный (Valid).

Правильно построенный документ соответствует всем правилам синтаксиса XML.

Действительный документ дополнительно соответствует некоторым семантическим правилам. Это более строгая дополнительная проверка корректности документа на соответствие заранее определенным, но уже внешним правилам в целях минимизации числа ошибок, например, структуры и состава данного, конкретного документа или

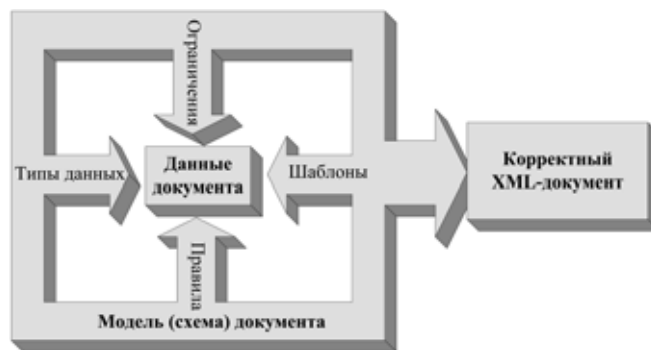


Рис. 1. Схема валидации XML-документа

семейства документов. Обычно такие правила хранятся в специальных файлах-схемах [5], где описана модель данных документа.

Рассмотрим подробнее процесс валидации XML-документов, для этого рассмотрим схему, изображенную на рис. 1.

Из анализа схемы валидации можно сделать вывод, что использование схемы (модели документа) обеспечивает контроль корректности вводимых данных в XML-документ. Документ сохранится без выдачи сообщения об ошибке только при полном соответствии модели документа.

Поддержка квазиструктурированной модели обеспечивается возможностями технологии XML. При описании модели документа можно определять необязательные, повторяющиеся, а также полностью неограниченные атрибуты. Кроме того, модель может быть расширена добавлением в нее новых элементов.

Пример использования

Рассмотрим применение методики для документов "Руководство пользователей АРМ" (АРМ — автоматизированное рабочее место) из пакета документов рабочей документации.

Прежде всего необходимо разработать модель документа. Модель документа определяет, какие данные, в каком формате и в какой последовательности будут храниться в XML-документе. Структура документа разрабатывается в соответствии с ГОСТ 19.505—79 "Руководство оператора. Требования к содержанию и оформлению". Для элементов схемы может быть задан порядок следования элементов и возможность обязательного и не обязательного использования элементов.

Схема документа "Руководство пользователей АРМ" состоит из сегментов (рис. 2). Каждый сегмент отвечает за описание раздела документа, а он в свою очередь подразделяется на более мелкие сегменты и атрибуты.

Сегменты "abbr_list" и "user_guide" показаны на схеме штриховой рамкой. При описании доку-

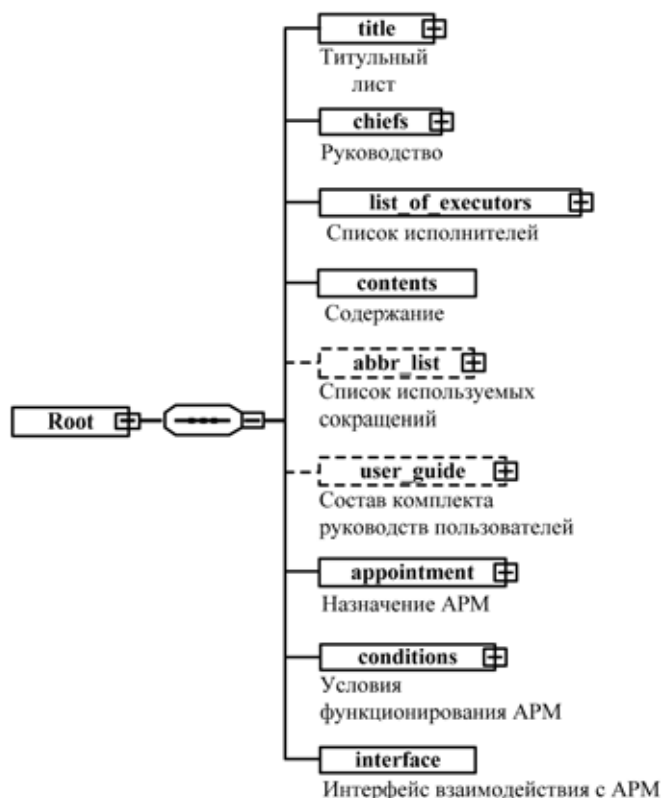


Рис. 2. Базовый сегмент схемы

мента эти элементы являются не обязательными и могут не присутствовать в документе.

Рассмотрим подробнее структуру модели.

Сегмент **"title"** описывает информационное наполнение титульного листа документа (рис. 3), которое состоит из атрибутов **"organization"** — организация-разработчик, **"project_name"** — название проекта, **"doc_name"** — название документа, **"doc_type"** — тип документа и двух сегментов: сегмента **"cipher"** — полный шифр документа, состоящего из двух атрибутов **"cipher_project"** — шифра проекта и **"cipher_doc"** — шифра документа, и сегмента **"dev_dt_plc"** — место и год создания документа, в котором определены два атрибута: **"dev_place"** и **"dev_date"** — место создания документа и дата создания документа соответственно.

На втором листе документа размещена информация о руководителях проектной организации, которая описана с помощью сегмента **"chiefs"** (рис. 4). Этот сегмент включает

в себя следующие атрибуты: **"engineer_institute"** — главный инженер института, **"engineer_project"** — главный инженер проекта, **"chief_department"** — начальник отдела. Каждый из перечисленных атрибутов будет содержать ФИО соответствующего руководителя.

Далее использован сегмент **"list_of_executors"** — список исполнителей документа (рис. 5). Этот сегмент состоит из списка авторов документа, для каждого из которых определены его ФИО и занимаемая им должность. Число исполнителей документа неограниченно.

Содержание документа описано с помощью атрибута **"contents"**, который представляет собой блок неограниченного размера.

Сегмент **"abbr_list"** позволяет описать сокращения, использованные в документе. Структура этого сегмента представлена на рис. 6. Число использованных аббревиатур в документе неограниченно. Каждое сокращение определено с помощью двух атрибутов: **"abbr_name"** — аббревиатура и **"abbr_decryption"** — расшифровка аббревиатуры.

Состав комплекта руководств пользователей описан с помощью сегмента **"user_guide"** (рис. 7). Число руководств комплекта неограниченно. Для каждого руководства определены следующие атрибуты: **"cipher_guide"** — шифр руководства и **"name_guide"** — название руководства.

Для описания функционального и эксплуатационного назначения АРМ используется сегмент **"apointment"** (рис. 8). Для него определены следующие атрибуты: **"functional_apointment"** — функциональное назначение, **"maintenance_apointment"** — эксплуатационное назначение, а также с помощью атрибута **"function_name"** — функции

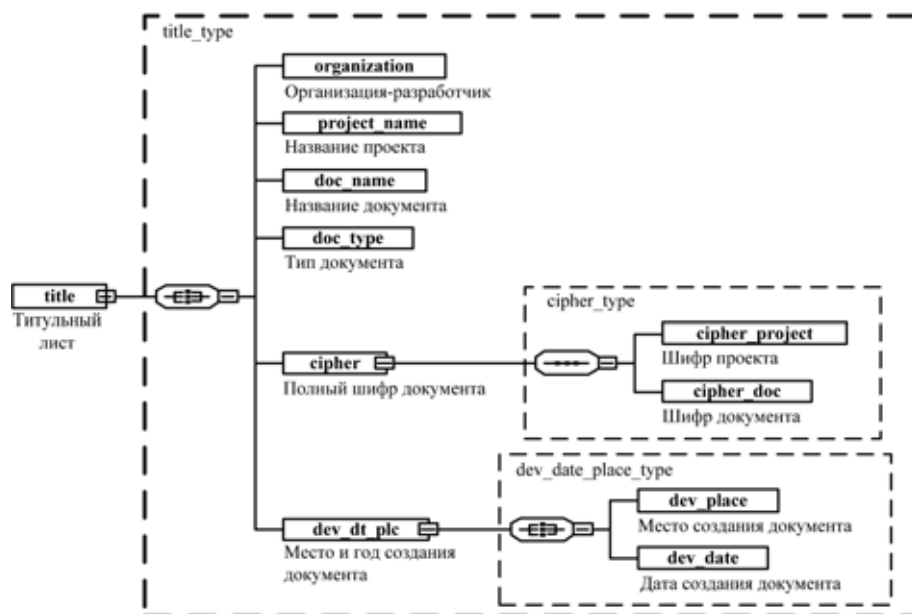


Рис. 3. Сегмент титульного листа

АРМ — может быть описано неограниченное число функций, реализуемых АРМом.

Далее в документе определены условия функционирования АРМ. Их описание реализует сегмент **"conditions"** (рис. 9), который в свою очередь состоит из сегментов **"min_system_requirements"** — минимальные системные требования, **"adv_system_**

requirements" — рекомендуемые системные требования и атрибутов, **"conditions_operator"** — условия эксплуатации АРМ, **"software"** — список программного обеспечения (ПО).

Атрибут **"conditions_operator"** определяет требования к рабочему месту оператора АРМ. Сегменты **"min_system_requirements"** и **"adv_system_requirements"** состоят из неограниченного числа атрибутов **"hardware_name"** — наименование и **"hardware_metries"** — характеристика. Сегмент, формирующий рекомендуемые системные требования, является необязательным и может отсутствовать в документе. Эти сегменты позволяют описать минимальные и рекомендованные требования к компьютеру АРМ. Неограниченная последовательность атрибутов **"software"** позволяет задать перечень программных средств, требуемых для функционирования АРМ на компьютере.

Атрибут **"interface"** использован в модели для маркирования разделов, содержащих описание интерфейса взаимодействия с АРМом.

Предложенная модель документов **"Руководство пользователей АРМ"** позволяет описать информационное наполнение документов этого вида. Использование предложенной модели позволяет упростить процесс создания документа, так как в этом случае документ разрабатывается в соответствии со структурой и информационным наполнением, определенным моделью. На рис. 10 изображен фрагмент документа, разработанный с использованием предложенной модели.

Документы, разработанные с использованием модели, позволяют реализовать доступ непосредственно к значимым атрибутам документа. В этом случае мы имеем дело с интерпретацией информационного наполнения документа на квазиструктурированное пространство данных [6].

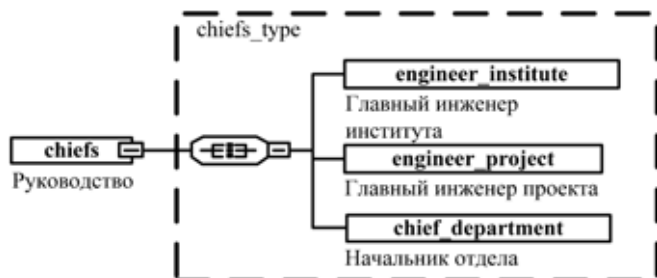


Рис. 4. Сегмент второго титульного листа

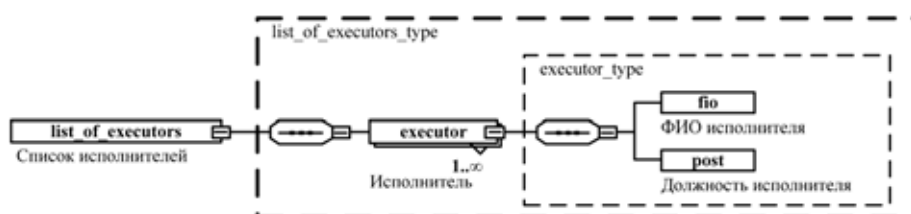


Рис. 5. Сегмент списка исполнителей документа

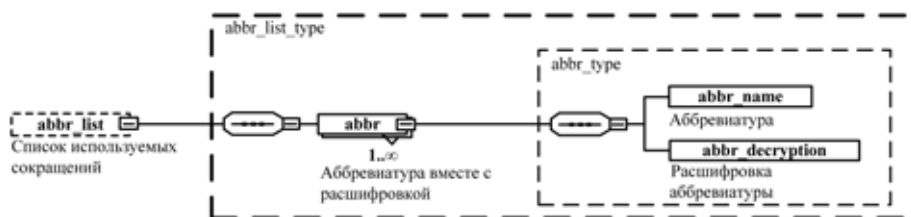


Рис. 6. Сегмент списка использованных сокращений

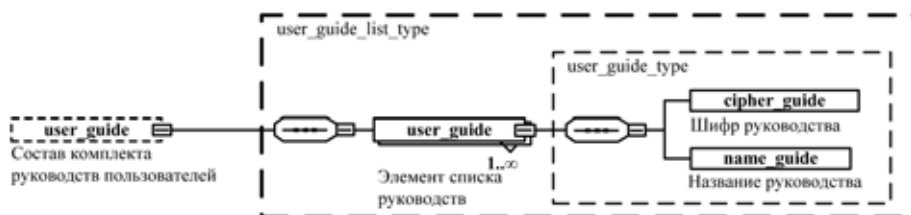


Рис. 7. Сегмент состава комплекта руководств пользователей

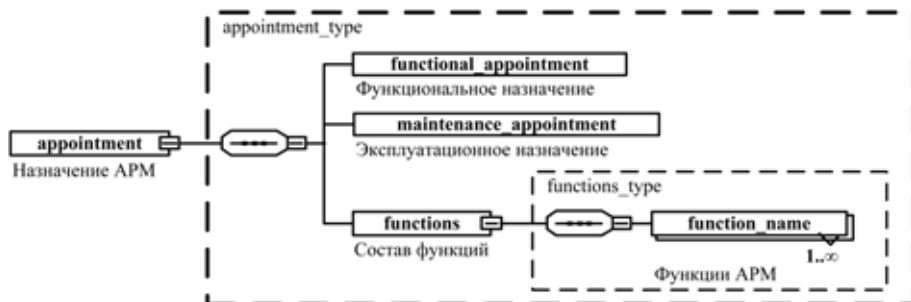


Рис. 8. Сегмент назначения АРМ

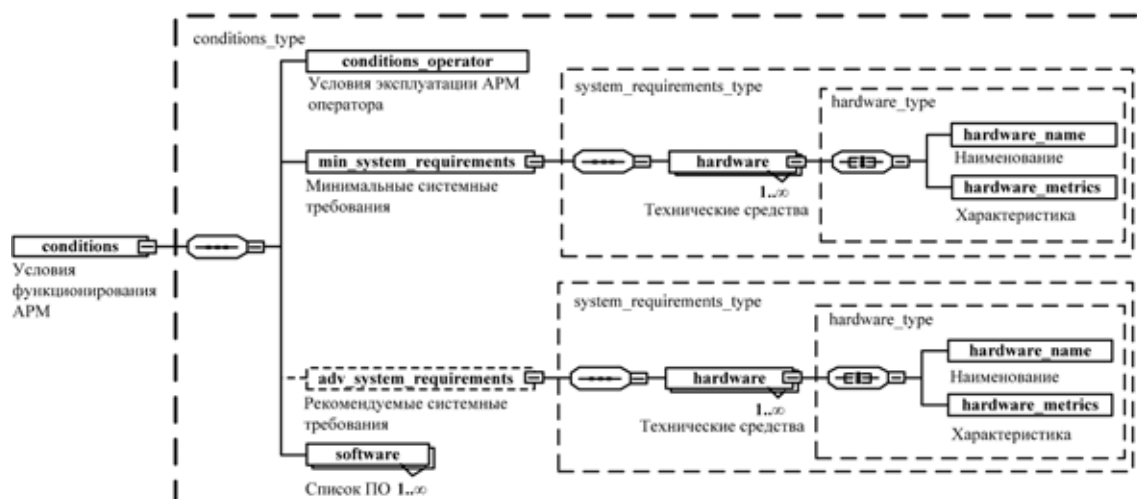


Рис. 9. Сегмент условия функционирования АРМ

Предложенный подход реализуется средствами СУБД Oracle 9i и более современными версиями [7].

Из документов, подготовленных по описанной модели, в автоматизированном режиме может быть получена:

- информация об организациях-разработчиках документов, названиях проектов и документов, типы документации, шифры проектов и документов, даты и место создания документов, ин-

формация о разработчиках документов включая информацию о руководителях проекта;

- содержание, список использованных сокращений (если присутствует) документа, состав комплекта руководств пользователей (если присутствует);
- информация о функциональном и эксплуатационном назначении АРМ, составе реализуемых им функций, условиях эксплуатации, ми-

Функциональное назначение

«functional_appointment» Функциональным назначением АРМ оператора БПТО и К является предоставление оператору возможности автоматизированного управления системой учета и транспортировки метанола. functional_appointment»

Эксплуатационное назначение

«maintenance_appointment» АРМ оператора БПТО и К предназначен для эксплуатации на технологическом объекте БПТО и К. Конечными пользователями АРМ являются операторы БПТО и К. maintenance_appointment»

Состав функций

АРМ оператора обеспечивает возможность выполнения следующих функций:

«functions» - function_name контроль давления метанола на входе в БСМ»;

- function_name контроль расхода метанола на выходе из БПТО и К»;

- function_name контроль давления метанола на выходе из БПТО и К»;

- function_name контроль температуры метанола на выходе из БПТО и К»;

- function_name контроль плотности метанола на выходе в БПТО и К»;

- function_name контроль массового расхода метанола на выходе в БПТО и К»;

- function_name сигнализация превышения плотности метанола на БПТО и К»;

- function_name включение/отключение режима опрессовки метанолопровода»;

- function_name сигнализация понижения давления в метанолопроводе».

functions» appointment»

Изм.	Кол.	Лист	№ док.	Подпись	Дата	07.082.088.3876-АСУ ИЗ.1	Лист
							9

Рис. 10. Фрагмент документа с маркированным информационным наполнением

нимальных и рекомендуемых системных требованиях (если присутствуют) и комплекте необходимого программного обеспечения.

Доступ к перечисленному информационному наполнению документов позволяет автоматизировано решать широкий круг задач. Например, для определенного технологического объекта может быть получен список программного обеспечения, необходимого для его функционирования. Этот список может быть использован для планирования процедуры его обновления.

Заключение

Предлагаемый в работе подход позволяет разрабатывать хранилища данных, совмещающие в себе преимущества документальных и фактографических систем. Кроме того в процессе наполнения таких хранилищ данных обеспечивается дополнительный контроль корректности информационного наполнения, который достигнут за

счет дополнительных ограничений, накладываемых моделями хранимых документов.

Список литературы

1. Корнеев В. В., Гареев А. Ф., Васютин С. В., Райх В. В. Базы данных. Интеллектуальная обработка информации. М.: Нолидж, 2000. 352 с.
2. Тенцер А. База данных — хранилище объектов // КомпьютерПресс. 2001. № 8. С. 144—145.
3. Палей Д. Моделирование квазиструктурированных данных // Открытые системы. 2002. № 09. С. 57—64.
4. Extensible Markup Language (XML) 1.0 (Fourth Edition) W3C Recommendation 16 August 2006. <http://www.w3.org/TR/REC-xml/>
5. XML Schema Part 1: Structures (Second Edition) W3C Recommendation 28 October 2004. <http://www.w3.org/TR/xmlschema-1/>
6. Полищук Ю. В. Интерпретация каталога документов на многомерное квазиструктурированное пространство данных // Основные проблемы освоения и обустройства нефтегазовых месторождений и пути их решения: тезисы докладов научно-технической конференции с международным участием. Оренбург, 26—27 июня 2008. С. 30—31.
7. Бен Чанг, Марк Скардина, Стефан Киритцов. Oracle9i XML. Разработка приложений электронной коммерции с использованием технологии XML. М.: Лори, 2003. 476 с.

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ И УПРАВЛЕНИЕ РИСКАМИ

УДК 621.395

М. А. Дрюченко, аспирант,

А. А. Сирота, д-р техн. наук, проф.,

Воронежский государственный университет, e-mail: aldrum@box.vsi.ru

Математическое и программное обеспечение для решения задачи стегоанализа на основе статистических и нейросетевых алгоритмов обработки информации

Описывается программный комплекс, предназначенный для анализа файлов различных форматов на предмет наличия в них информации, встроенной различными способами стеганографического скрываются. Рассматриваются возможности комбинированного использования нейросетевых и статистических алгоритмов анализа данных в интересах обнаружения стеганографически скрытой информации.

Ключевые слова: стеганография, стегоанализ, статистические алгоритмы, нейронная сеть, разладка, программный комплекс

Введение

Компьютерная стеганография относится к числу современных эффективных методов по защите конфиденциальной информации, хранению и передаче данных. Защищаемые данные скрытно встраиваются в структуру других, называемых стегоконтейнерами (СК) [1]. Существует достаточно много алгоритмов стеганографического скрываются информации, учитывающих как особенности форматов СК, так и особенности человеческого восприятия (психовизуальная избыточность СК). Наиболее распространены алгоритмы скрываются,

тем или иным способом модифицирующие наименее значимые биты (НЗБ) СК как в пространственной области, так и в области преобразования (LSB — *least significant bit methods*).

Проблемой выявления скрытой информации занимается стегоанализ (СА) [1]. Одно из перспективных направлений СА заключается в создании алгоритмов, позволяющих анализировать информацию без привязки к конкретным программам скрываются и реализованным в них алгоритмам. Важной задачей СА является определение существенных свойств СК, выявление их оп-

ределенных признаков, которые удовлетворяли бы требованиям репрезентативности и контекстной независимости и одновременно менялись бы при встраивании в контейнер дополнительной информации. Способность к противодействию СА определяет стойкость [1] стеганографических систем. Для обнаружения изменения свойств СК можно использовать искусственные нейронные сети (ИНС) [2, 3]. Аппарат ИНС позволяет достаточно эффективно восстанавливать характер эталонного СК, использованного для обучения сети, и одновременно отслеживать даже незначительные изменения различных статистических характеристик и корреляционных зависимостей в СК при появлении в нем аномальных изменений. Далее предлагается использовать данный факт в интересах решения задачи СА.

Цель данной работы — разработка математического и программного обеспечения, реализующего различные статистические и нейросетевые методы и алгоритмы для выявления фактов СС в файлах различных форматов.

Нейросетевой алгоритм выявления стеганографического скрывания данных с использованием методов обнаружения "разладки"

Под началом области стеганографического скрывания (СС) будем понимать определенное изменение свойств анализируемых последовательностей данных как в рамках последовательности СК, так и в рамках одного СК. В качестве примера СК будем рассматривать аудиофайлы формата WAV. Пусть $\{x_k\}$ — отсчеты звукового СК или других наблюдений, представляющие последовательность случайных величин (векторов) с условной плотностью вероятности $P_\theta(x)$, где $\theta \in R^r$ — некий вектор параметров. Можно предположить, что в результате стеганографического встраивания данных параметр θ меняет свое значение, начиная с некоторого неизвестного момента времени t_0 , что в свою очередь приведет к возникновению эффекта "разладки" [4]. Иными словами, до момента t_0 выполняется $\theta = \theta_0$, а после $\theta = \theta_1$, $\theta_0 \neq \theta_1$. Таким образом, задача поиска разладки, а значит, и задача отыскания начала области СС состоит в определении изменений параметра θ и определении момента времени t_0 .

Для обнаружения момента разладки, как правило, используются последовательные критерии, фиксирующие изменчивость свойств статистических оценок временных рядов на основе некоторых характеристик. Одним из наиболее эффективных критериев является алгоритм кумулятивных сумм (АКС) [4], реализующий вычисление статистики вида

$$g_t = (g_{t-1} + \Delta g_t)^+; \Delta g_t = \ln \left(\frac{\omega(x_t^{(n)}/\theta_1)}{\omega(x_t^{(n)}/\theta_0)} \right), \quad (1)$$

где $[x]^+ = \max[0, x]$; g_t — кумулятивная сумма на шаге t ; $\omega(x/\theta_0)$ — плотность распределения последовательности анализируемых наблюдений до момента наступления разладки; $\omega(x/\theta_1)$ — плотность распределения последовательности отсчетов СК после наступления разладки.

Фиксация разладки осуществляется при превышении кумулятивной суммой заданного порога $g_t > h$. Порог h целесообразно определять растущим со временем t , например, в виде

$$h(t) = C + \ln(t) + 2\ln(\ln(t)), \quad t = 1, 2, \dots, N, \quad (2)$$

где C — константа, отвечающая за вероятность ложной тревоги [4].

Если значения θ_0 и θ_1 до и после изменения свойств процесса известны неточно, но известно направление скачка в пространстве параметров, определяемое вектором $c = (c_1, \dots, c_r)^T$, может быть использовано асимптотическое разложение Ле Кама [4] для логарифма функции правдоподобия. При этом АКС может быть записан в виде

$$g_t = \left[g_{t-1} + \sum_{j=1}^n \sum_{q=1}^r \frac{\partial \ln \omega(x_j^{(t)}/\theta_0)}{\partial \theta_{0q}} c_q \right]^+, \quad (3)$$

$g_0 = 0$; $\tilde{t}_0 = t_\alpha - n_\alpha + 1$; h — порог принятия решения о возникновении разладки; t_α — момент подачи сообщения о разладке; n_α — число шагов, выполненных от момента последнего обнуления кумулятивной суммы g_t до момента принятия решения.

Схема работы предлагаемого нейросетевого алгоритма выявления СС включает реализацию двух этапов:

- формирование авторегрессионной модели (АМ) анализируемых процессов в стационарном режиме (когда происходит обучение нейросети (НС) на "чистых" СК);
- тестирование НС и поиск разладок на основе наблюдения реакции сети (при отсутствии информации о наличии стегоскрывания).

При формировании обучающих множеств и при тестировании НС осуществляется предварительная обработка звуковых файлов-контейнеров. В целях уменьшения нестационарности из исходного сигнала $S(t)$ (рис. 1, а) исключаются единичные скачки и участки, содержащие серии отсчетов с одинаковыми значениями. Далее сигнал подвергается высокочастотной дискретной фильтрации, и полученные значения нормируются от -1 до 1 . В результате получаем новый сигнал $\tilde{S}(t)$ (рис. 1, б).

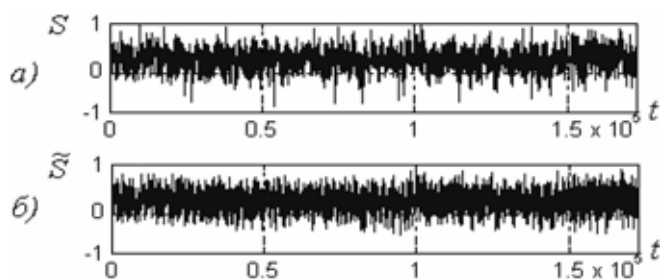


Рис. 1. Исходный звуковой сигнал (а); отфильтрованный звуковой сигнал (б). По оси абсцисс — дискретное время, по оси ординат — значения отсчетов сигналов, нормированные от -1 до 1

На следующем шаге формируются так называемые матрицы "входных воздействий" нейронной сети. Обозначим $\tilde{S} = (\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_N)^T$ — вектор, содержащий последовательность отсчетов обработанного звукового сигнала. Тогда матрица "входных воздействий" будет определяться как

$$Y = \{y_1, y_2, \dots, y_{N-\delta}\},$$

где $y_i = (\tilde{s}_i, \tilde{s}_{i+1}, \dots, \tilde{s}_{i+\delta-1})$, $i = 1, 2, \dots, N - \delta$; δ — длина вектора y_i . Значения векторов y_i далее подаются на входы ИНС.

По своей архитектуре НС представляет собой сеть класса многослойный перцептрон и может быть как линейной, так и нелинейной. Число скрытых слоев, число нейронов в скрытых слоях, а также функции активации определяются пользователем на этапе конструирования сети. Величины, получаемые в простейшем случае на выходе линейной сети, идентифицируются уравнением авторегрессии вида

$$\tilde{y}_t = \sum_{i=1}^{\delta} W_i y_{t-i} + A + \eta_t, \quad (4)$$

где $t = \delta, \delta + 1, \dots, N$; W_i — матрицы, составленные из весов НС, обеспечивающие прохождение входных воздействий y_{t-i} на выход; A — вектор смещений нейронов сети; η_t — вектор ошибки предсказания значений $\tilde{y}_t$ на выходе НС, имеющий математическое ожидание, равное нулю, и неизвестную матрицу ковариации R_{η} .

При построении АМ модели СК рассматривалось несколько вариантов формирования компонент матриц входных воздействий: *предсказание одного отсчета наблюдаемых данных в каждый момент времени, векторное предсказание внутри и на границе вектора отсчетов звукового сигнала.*

Параметры W_i , $i = \overline{1, \delta}$ и A из (4) определяются в ходе обучения НС. Оценка матрицы ковариации вектора η_t АМ и математическое ожидание среднеквадратического отклонения (СКО) проводятся в ходе тестирования сети после обучения на тех же данных [3]:

$$R_{\eta} = \frac{1}{N} \sum_{t=1}^N (y_t - \tilde{y}_t)(y_t - \tilde{y}_t)^T, \quad (5)$$

где y_t — вектор из обучающей (тестирующей) выборки данных; $\tilde{y}_t$ — предсказываемый сетью вектор. В качестве компонентов вектора параметров θ АМ модели, используемого для обнаружения разладки, рассматривается СКО выхода НС.

В ходе исследований были рассмотрены несколько моделей данных, подаваемых на входы ИНС, которые схематично представлены на рис. 2. На рис. 2, а представлен рассмотренный выше случай, когда компоненты матрицы входных воздействий Y формируются только из отсчетов анализируемого сигнала. На рис. 2, б приводится схема формирования компонент матрицы Y вторичными данными, полученными после применения различных статистических стегоаналитических критериев по отношению к исходному СК. В этом случае разладка уже ищется по отношению к последовательности результатов обработки данных с помощью различных статистических алгоритмов СА, применяемых к СК. На рис. 2, в представлен случай совместного формирования вектора входных воздействий НС на основе как отсчетов анализируемого сигнала, так и результатов применения к СК статистических алгоритмов обработки первичных данных в виде

$$Y = \{y_1, y_2, \dots, y_{N-\delta}, M_1, M_2, \dots, M_k\},$$

где M_i , $i = 1, 2, \dots, k$ — значения, полученные в результате применения к СК статистических методов анализа. Величины M_i здесь выступают в роли сопутствующих или управляющих параметров (УП) НС.

Практическая реализация процедуры СА включает несколько шагов. Сначала проводится обучение НС в стационарном режиме (верхние схемы рис. 2, а—в) при использовании реализаций СП без скрытой информации. При этом пользователь загружает обученную ИНС из файла либо создает новую сеть необходимой конфигурации и обучает ее на тестовых данных, сформированных из значений отсчетов звуковых файлов, не содержащих СС информации.

В процессе тестирования НС проводится настройка параметров системы на вероятность ложного обнаружения разладки, определяется порог принятия решений $h(t)$. Далее на входы обученной НС подаются векторы входных воздействий, сформированные согласно одной из схем рис. 2. На выходе сети имеем предсказанные значения $\tilde{y}_t$. Параллельно в каждый момент времени t вычисляется СКО предсказания НС очередного вектора y_t

$$Z_t^{(1)} = z_t = \frac{1}{r} \sum_{i=1}^r (y_t^{(i)} - \tilde{y}_t^{(i)})^2, \quad (6)$$

где $y_t^{(i)}$ — вектор значений, подаваемый на входы НС; $\tilde{y}_t^{(i)}$ — вектор, полученный на выходе сети.



Рис. 2. Схемы обучения и тестирования нейронных сетей для различных типов входных данных

В программе реализованы два вида решающих правил, настроенных на определение скачка среднего значения СКО выхода НС. Первое получено на основе применения формулы разложения Ле Кама (3) в случае, когда вектор параметров формируется на основе первого начального момента СКО выхода НС

$$\frac{\partial \ln \omega(z_t/\theta)}{\partial \ln \theta} \Big|_{\theta=\theta_0} = \frac{z_t - m_z \xi}{\sigma_z};$$

$$g_t = \left[g_{t-1} + \frac{z_t - m_z \xi}{\sigma_z} \right]^+, \quad (7)$$

где $\theta = m_z$; $t_\alpha = \inf\{t: g_t \geq h\}$; $g_0 = 0$; $\tilde{t}_0 = t_\alpha - n_\alpha + 1$; $0 < \xi < 2$ — коэффициент "чувствительности" к ложным тревогам; m_z и σ_z — математическое ожидание и дисперсия СКО, которые оцениваются в ходе обучения НС.

Если предположить, что в результате СС в отсчетах звукового файла изменится вид закона распределения в соответствующих битовых плоскостях, то должны измениться как m_z , так и σ_z , т. е. возникнет смещение $m_z \rightarrow m'_z$ и $\sigma_z \rightarrow \sigma'_z$. В этом случае вектор параметров θ формируется из начальных моментов первого и второго порядка СКО выхода НС $\theta = (m_z, \sigma_z)^T$, и второе решающее правило можно записать в виде

$$g_t = \left[g_{t-1} + c_1 \frac{(z_t - m_z^{(0)} \xi)}{(\sigma_z^{(0)})^2} + c_2 \frac{(z_t - m_z^{(0)} \xi)^2 - (\sigma_z^{(0)})^2}{(\sigma_z^{(0)})^3} \right]^+; \quad (8)$$

$$c_1 = \frac{(m'_z - m_z)}{\sqrt{(m'_z - m_z)^2 + (\sigma'_z - \sigma_z)^2}},$$

$$c_2 = \frac{(\sigma'_z - \sigma_z)}{\sqrt{(m'_z - m_z)^2 + (\sigma'_z - \sigma_z)^2}},$$

где $g_0 = 1$; ξ — параметр чувствительности алгоритма к ложным тревогам, $\tilde{t}_0 = t_\alpha - n_\alpha + 1$; $\theta_0 = (0,5(m_z + m'_z), 0,5(\sigma_z + \sigma'_z))^T$; c_1, c_2 — компоненты вектора направления разладки для параметров m_z и σ_z соответственно.

На рис. 3, б, в в логарифмическом масштабе приведены графики изменения кумулятивных сумм, фиксирующие моменты возникновения разладки в форме (7) и (8).

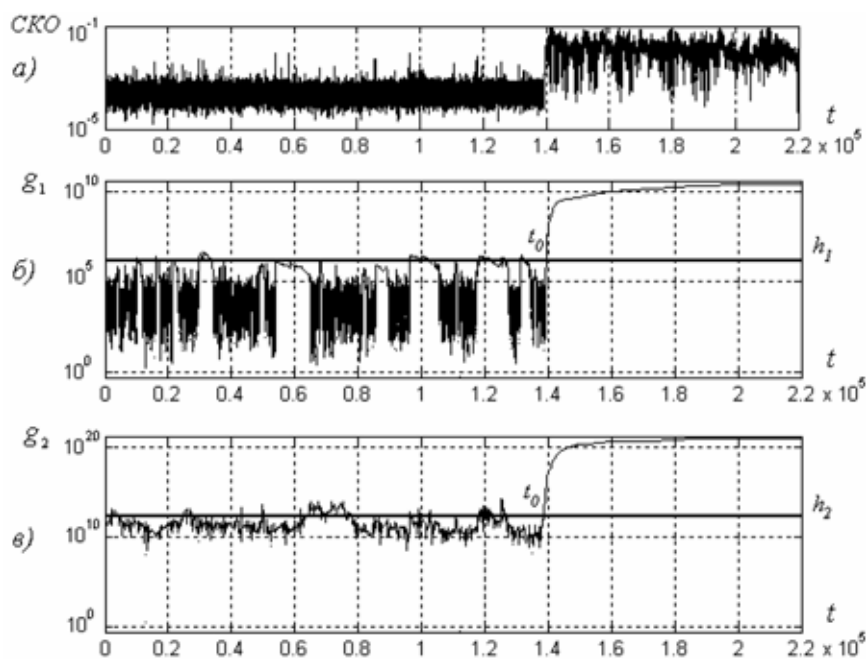


Рис. 3. Временная диаграмма СКО восстановления звукового сигнала нейронной сетью (а); графики кумулятивных сумм, фиксирующие момент возникновения разладки (б, в)

Как показали результаты экспериментов, применение решающего правила в форме (8) позволяет более точно (с меньшим запаздыванием) фиксировать момент обнаружения разладки, однако число ложных тревог оказалось выше, чем при использовании АКС в форме (7).

Статистические алгоритмы стегоанализа

В рамках рассматриваемого подхода в работе был реализован ряд статистических алгоритмов стегоанализа, подключаемых, как было описано ранее, к ИНС в различных режимах. Некоторые из них являются универсальными и могут с незначительными изменениями применяться для анализа файлов различных типов. Другие алгоритмы стати-

Таблица 1

Статистические критерии, используемые при анализе звуковых файлов

Статистические методы СА	Краткая характеристика
Метод оценки числа переходов значений в потоке наименее значимых бит СК	При СС данных распределение наименее значимых бит СК зачастую приобретает случайный характер. Метод отслеживает изменение вида закона распределения младших бит СК
Метод оценки статистики длин битовых серий	Длинные серии одинаковых битовых значений более характерны для "чистых" СК, в процессе СС данных уменьшается число длинных и увеличивается число серий единичной длины
Критерий согласия хи-квадрат	Оценивается соответствие реально наблюдаемой статистики ожидаемой. Проверка осуществляется на уровне монобитов, дибитов

стического СА используют особенности конкретных форматов. Например, при поиске стеговложений в JPEG-файлах в программе анализируются гистограммы квантованных дискретных косинусных коэффициентов, анализируется число ошибок округления значений цветовых составляющих при декомпрессии изображения [5] и т. д. Учитывая, что невозможно предусмотреть все возможные варианты СС, задача сводится к созданию стегоаналитической системы, реализующей максимально широкий набор методов СА и комплексное их использование с тем, чтобы минимизировать ошибки первого и второго рода при классификации СК.

В табл. 1 приведены статистические критерии, используемые при анализе звуковых файлов на наличие в них СС информации.

Описание разработанного программного комплекса

Программное обеспечение для решения задач стегоанализа на основе различных алгоритмов обработки информации разработано в среде "Borland C++ Builder 6" для работы в операционной системе Windows 2000/Windows XP. В качестве языка разработки был выбран C++ как один из наиболее универсальных и легко переносимых на другие платформы.

Обобщенная структура программного комплекса для встраивания и выявления СС информации представлена на рис. 4. На ней приведены наиболее важные программные модули, сгруппированные по своему назначению: реализующие СС данных и реализующие комплекс методов для СА.

Разработанные модули для СА позволяют анализировать файлы форматов BMP, GIF, PNG, JPEG, WAV. На рис. 5, в приведено главное окно приложения с активной вкладкой "Стегоанализ". На ней пользователь загружает файлы для анализа. Если их формат известен программе, то применяются стегоаналитические методы (статистические и нейросетевые), специфичные для данного формата файлов. В результате выводится набор графиков, гистограмм и оценок, по которым пользователь может судить о вероятности наличия скрытых вложений. В программной части для выявления СС данных также реализованы методы "сигнатурного" анализа (когда в специальных областях файлов ищутся битовые последова-

тельности, специфичные для определенных программ стегоскрытия).

Кроме модулей, предназначенных для решения задач СА, настоящий программный комплекс содержит ряд модулей, в которых реализованы процедуры стеганографического скрытия. Модули СС позволяют встраивать информацию в файлы форматов BMP, GIF, PNG, JPEG, txt, html, mp3, WAV. В файлы, предусматривающие сжатие, встраивание происходит в область преобразования, например, в JPEG-файлы скрытие проводится в квантованные дискретные косинусные коэффициенты. При этом скрываться могут файлы любых форматов. Перед встраиванием они дополнительно сжимаются по методу LZSS (LZ — Lempel-Ziv, S — Storer, S — Szimanski). В программе также предусмотрено шифрование встраиваемых данных по алгоритму ГОСТ 28147—89.

Размеры СК не ограничиваются. Например, для скрытия информации успешно использовались JPEG-контейнеры размерами от нескольких килобайт до 30—60 Мбайт. При этом была проде-

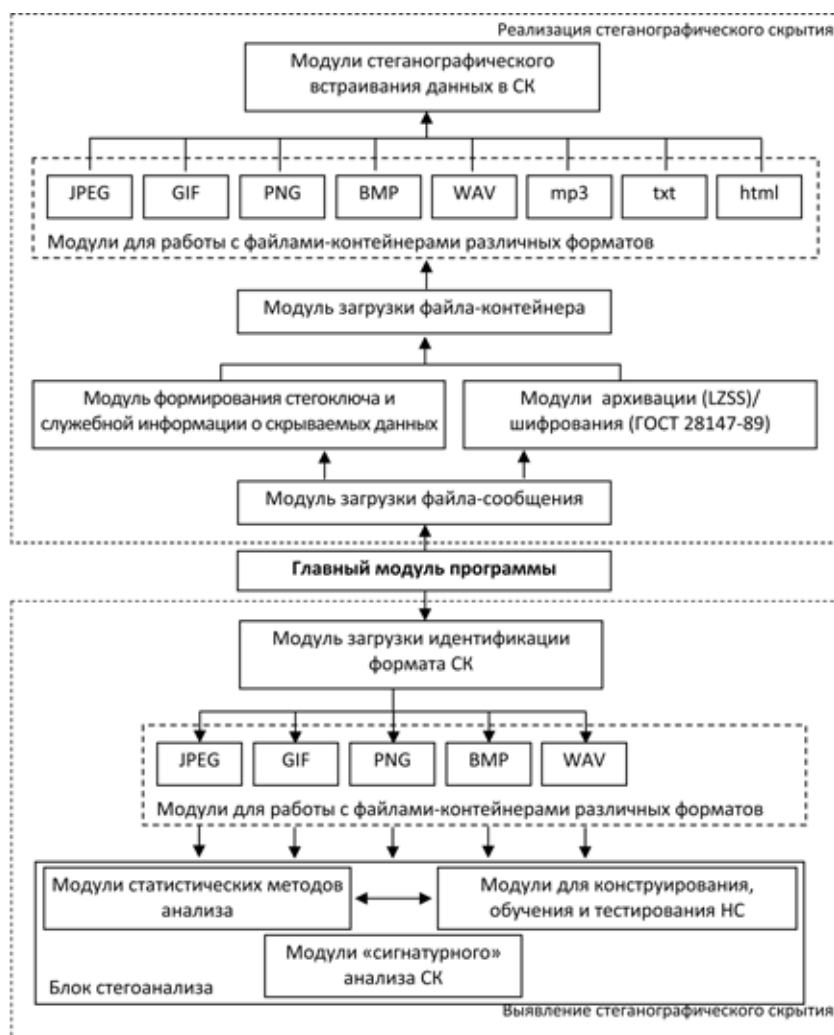


Рис. 4. Структура программного комплекса для стеганографического встраивания и выявления СС информации

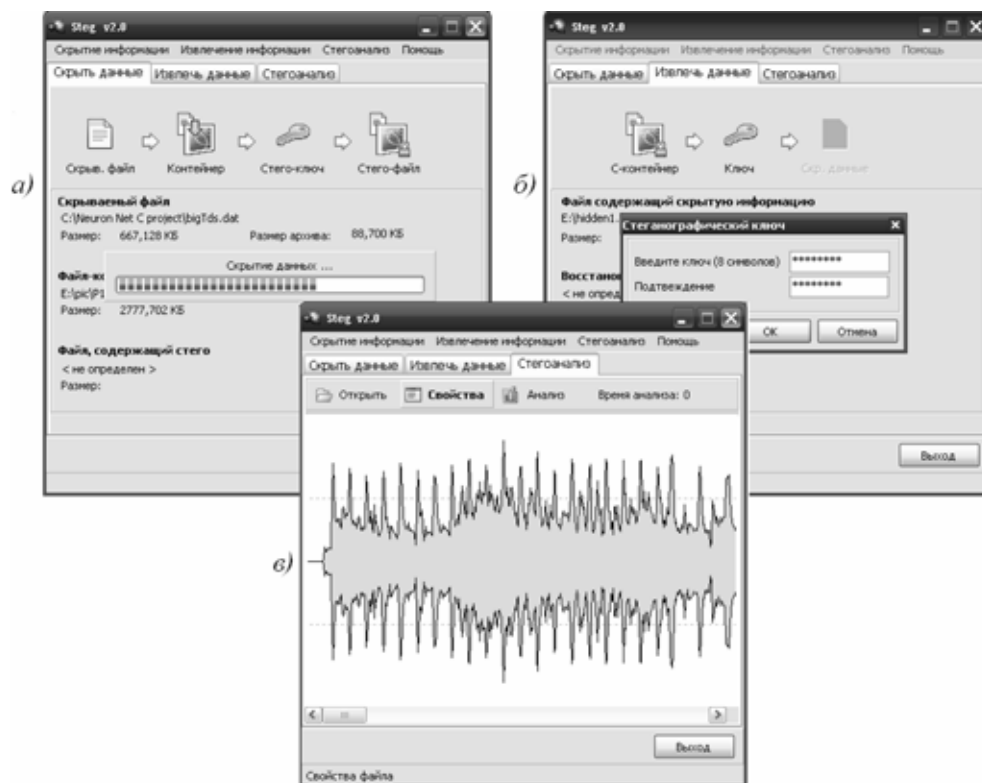


Рис. 5. Главное окно программы стеганографического скрытия/анализа данных:
а — вкладка "Скрыть данные"; б — вкладка "Извлечь данные"; в — вкладка "Стегоанализ"

монстрирована достаточно высокая скорость работы. На рис. 5, а и б представлено главное окно программы с активными вкладками "Скрытие данных" и "Извлечение данных".

Представленный в работе программный комплекс СС/СА является в большей степени исследовательским и прежде всего предназначен для отработки различных технологий стегоанализа. Несмотря на это по своим возможностям он сходен, а по многим параметрам и превосходит, многие известные программы СС и СА данных.

Библиотека классов для работы с ИНС

В работе была отдельно реализована библиотека классов на языке C++ для работы с ИНС. Следует отметить, что, безусловно, самым распространенным программным пакетом, предоставляющим возможность работы с НС, является Neural Network Toolbox системы MatLab. Основное преимущество данного пакета заключается в возможности моделирования очень широкого класса НС, а его основные недостатки — это громоздкая реализация и невысокая скорость работы при решении задач определенного класса. Кроме того, использование нейросетевых средств MatLab накладывает значительные ограничения на переносимость разрабатываемой программы СА. Поэтому разумным шагом является отдельная про-

граммная реализация модулей для работы с НС (максимально оптимизированная для решения задач СА) и их интеграция с основными модулями разрабатываемой стегоаналитической системы.

Функциональные возможности разработанных программных модулей для работы с НС включают:

- возможность создания многослойных НС необходимой конфигурации с указанием числа слоев, числа нейронов в каждом слое и вида активационных функций (линейная, сигмоидная или гиперболический тангенс) (рис. 6, а);
- несколько вариантов начальной инициализации весовых коэффициентов ИНС (рис. 6, б);
- несколько методов обучения многослой-

ных сетей, среди которых *метод градиентного спуска*, *метод сопряженных градиентов*, *квазиНьютоновский метод* и несколько модификаций *метода "неунывающего" обратного распространения RPROP (Resilient backpropagation)*.

Настройка параметров обучения, таких как точность обучения, число эпох и т. д. осуществляется на вкладке "Обучение" (рис. 6, в). В процессе обучения отображаются графики ошибки, нормы градиента и значений улучшения решения (рис. 6, г). Предусмотрена возможность тестирования сети после ее обучения. Также реализована возможность сохранения и загрузки обученной НС из файла.

Реализованные классы для работы с НС имеют самостоятельное значение и могут применяться как в рамках процедуры СА (см. рис. 2), так и отдельно. Их можно применять для решения разного рода задач без необходимости установки специализированных пакетов (таких как MatLab).

Экспериментальные исследования

При проведении экспериментальных исследований были проанализированы результаты работы ряда широко известных стеганографических утилит, осуществляющих скрытие в различные типы файлов. Среди программ, реализующих СС в WAV-файлы, рассматривались *S-Tools*, *Stegomagic*, *Steganos for Windows*, *Steganos Security Suite*, *Fox Secret*. Тестиро-

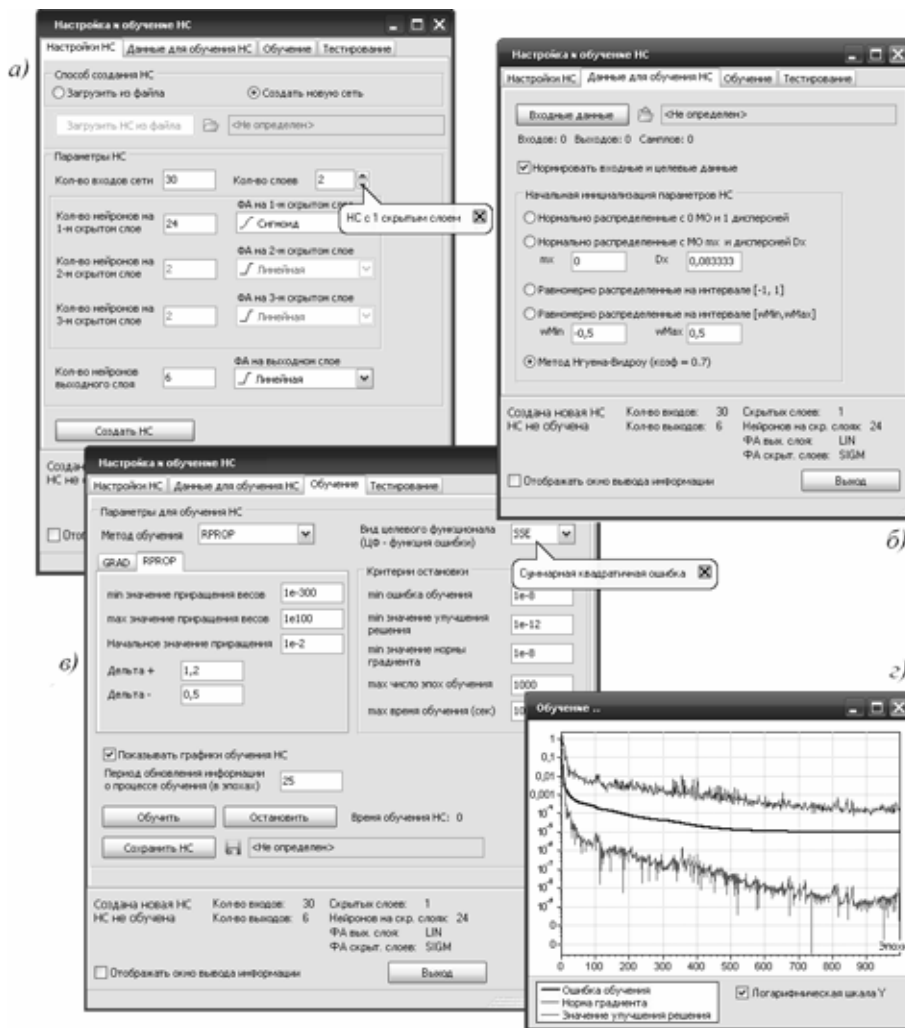


Рис. 6. Графический интерфейс программных модулей для работы с ИНС:
 а — вкладка создания и настройки конфигурации нейронной сети; б — вкладка загрузки обучающих данных; в — вкладка обучения сети; г — окно вывода графиков ошибки обучения, нормы градиента и улучшения решения

Таблица 2

Результаты применения нейросетевого алгоритма выявления стеганографического скрывания

Программы стеганографического скрывания	S-Tools	Stego magic	Steganos for Windows	Fox Secret	Sreganos Security Suite
Достоверное обнаружение стегоскрывания, %	82,4	92	86	90	77

вание методов проводилось на специально сформированном наборе из 108 звуковых файлов, в которые проводилось скрывание информации с использованием перечисленных стеганографических утилит. Некоторые из результатов тестирования для звуковых образцов, на 50 % и более заполненных встраиваемой информацией, приведены в табл. 2. В данном

случае векторы входных воздействий ИНС формировались из отсчетов анализируемого сигнала и результатов применения к СК статистических алгоритмов, приведенных в табл. 1, а для определения момента разладки, т. е. начала области СС, использовалось решающее правило в форме (7).

На результаты работы алгоритма в значительной степени влияют:

- отношение длины скрываемого сообщения и длины СК;
- алгоритм, по которому происходит встраивание информации;
- число байтовых разрядов, подвергшихся модификации;
- характеристики самого СК, такие как отношение сигнал/шум и наличие в контейнере областей существенной нестационарности.

Запаздывание обнаружения разладок оценивалось исходя из формулы

$$\tilde{\tau} = t_{\text{скр}} - t_0 + 1,$$

где $t_{\text{скр}}$ — известный момент начала области СС; t_0 — момент обнаружения разладки. В случае отсутствия ошибок 1-го и 2-го рода значение запаздывания составляло $280 < \tilde{\tau} < 1900$ отсчетов сигнала.

Тестирование разработанных программных модулей СА показало их высокую эффективность, а комплексное применение различных статистических и нейросетевых алгоритмов СА позволяет более гибко подходить к вопросу о выявлении скрытно встроенной информации, а также обеспечивать наглядность результатов анализа.

Список литературы

1. Грибунин В. Г. Цифровая стеганография. СПб.: СОЛОН-Пресс, 2002. 280 с.
2. Bergman C., Davidson J. An Artificial Neural Network for Wavelet Steganalysis, October 1, 2004—September 31, 2006. (www.mfrc.ameslab.gov/R-D_Reports/Artificial_Neural_Network_for_Steganalysis_Final.pdf).
3. Дрюченко М. А., Сирота А. А. Нейросетевой алгоритм выявления стеганографически скрытых сообщений методом обнаружения "разладки" // Радиотехника. 2008. № 11. С. 24—28.
4. Никифоров И. В. Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1984. 199 с.
5. Дрюченко М. А. Алгоритмы выявления стеганографического скрывания информации в JPEG-файлах // Вестник Воронежского государственного университета. 2007. № 1. С. 21—31.

Г. А. Кустов, канд. техн. наук,
ведущий программист,

Уфимский филиал Центрального
коммерческого банка,

М. А. Николаева, канд. техн. наук, доц.,

О. Ф. Зотова, канд. техн. наук, доц.,

Р. А. Шарапов, аспирант,

Уфимский государственный авиационный
технический университет,
e-mail: olgazotova@list.ru

Алгоритмы выбора превентивных мер снижения рисков

Рассмотрен подход к анализу, оценке и управлению рисками на примере информационных рисков. Основу этого подхода составляет логико-вероятностный метод. Предварительный анализ информационных рисков, использующий карту рисков, позволил выделить наиболее значимые опасные состояния ресурсов, риск которых не игнорируется и не передается на управление, а снижается за счет внедрения превентивных мер — контрмер. Предложены алгоритмы формирования эффективного набора контрмер, выбор каждого из которых определяется исходной информацией. При этом алгоритмы являются следствием применения логико-вероятностного метода для идентификации и оценки рисков.

Ключевые слова: анализ и оценка рисков, информационные риски, логико-вероятностный метод, опасные состояния, угрозы, значимость угрозы, контрмеры, алгоритм формирования эффективного набора контрмер

Введение

Задача управления рисками возникает в разных предметных областях, однако основные подходы к анализу, оценке и управлению можно считать сходными.

Формально решение задачи управления рисками (если говорить о снижении риска) сводится к выполнению следующих этапов:

- идентификация ресурсов системы. Примерами систем могут быть экосистемы, информационные системы организаций, финансовые системы и др.;
- идентификация и анализ угроз, действующих на эти ресурсы. Угрозы представляют собой производственные выбросы, программные и аппаратные отказы, механические сбои, невозможность выполнения обязательств и т. п.;
- оценка уровня рисков для каждого ресурса и системы в целом;

- выбор мер для снижения рисков (контрмер) и анализ эффективности управления.

Оценка уровня риска системы

Под *риском системы* понимается сумма рисков ресурсов, из которых она состоит:

$$R = \sum_{i=1}^n R_i, \quad (1)$$

где R_i — риск i -го ресурса; n — число ресурсов.

С каждым ресурсом связано множество опасных состояний (ОС), реализация которых приводит к отказу данного ресурса.

Под *риском i -го ресурса* понимается сумма рисков, связанных с реализацией опасных состояний данного ресурса:

$$R_i = \sum_{j=1}^{M_i} r_{ij}, \quad (2)$$

где r_{ij} — риск реализации j -го опасного состояния i -го ресурса, $j = \overline{1, M_i}$; M_i — число опасных состояний i -го ресурса.

Под *риском реализации j -го опасного состояния i -го ресурса* понимается произведение вероятности P_{ij} и стоимости потерь C_{ij} от реализации данного опасного состояния ресурса:

$$r_{ij} = P_{ij} \cdot C_{ij}. \quad (3)$$

Таким образом, задачу оценки риска системы можно разбить на следующие этапы:

- описание структуры ресурсов системы;
- описание множества опасных состояний ресурсов системы;
- оценка вероятностей P_{ij} реализации опасных состояний, в том числе выявление меры влияния угроз на реализацию опасных состояний;
- оценка стоимости потерь C_{ij} от реализации опасных состояний.

Предварительная оценка рисков ресурсов

Важным этапом задачи оценки текущего уровня риска системы является классификация имеющихся рисков в целях выбора метода управления этими рисками.

Существует общепринятая классификация рисков *по частоте возникновения ущерба*, в соответствии с которой выделяются следующие классы рисков [1]:

- *редкие риски*, для которых характерна малая частота реализации риска, т. е. малая вероятность наступления ущерба;

Группировка рисков по частоте возникновения и размеру ущерба

По размеру	По частоте		
	Редкие	Средней частоты	Частые
Малые риски	—	—	+
Средние риски	+	+	+
Высокие риски	+	+	—
Катастрофические риски	—	—	—

- *риски средней частоты*, для которых характерна средняя частота реализации риска, т. е. средняя вероятность наступления ущерба;
- *частые риски*, для которых характерна высокая частота реализации риска, т. е. высокая вероятность наступления ущерба.

Политика управления рисками для указанных классов будет существенно различаться.

Другой способ классификации рисков — **по размеру ущерба** (потерь), выражаемого денежными единицами ущерба. Если имеется полная информация об угрозах, уязвимостях и стоимости ресурсов, то определение размера ущерба не вызывает никаких проблем. Если же такой информации недостаточно, то соответствующие классы рисков можно установить на основе экспертного заключения. По данному критерию можно выделить:

- *малые риски*, т. е. те, по которым максимальный ущерб невелик;
- *средние риски*, максимальный ущерб для которых характеризуется как средний;
- *высокие риски* с большим максимальным ущербом;
- *катастрофические риски*, характеризующиеся исключительно большим максимальным ущербом.

Подобная классификация чрезвычайно важна и широко используется на практике (см. таблицу).

Такая классификация позволяет понять специфику различных рисков. Очевидно, что для рисков, выделенных в соответствии с данной классификацией, методы анализа и управления будут совершенно разными.

Малые риски — редкие и со средней частотой — не требуют внимания андеррайтеров, т. е. управление рисков сводится к их *игнорированию, отказу от управления*. Если говорить о катастрофических рисках, то, конечно же, их надо страховать, т. е. осуществлять *передачу управления риском*. Если ресурс подвержен высоким и частым рискам, то его лучше просто заменить — *отказ от ресурса*. Остальные риски, отмеченные в таблице знаком плюс, требуют анализа, определенной степени детализации при оценке, управление ими сводится к осуществлению превентивных мер по *снижению риска*.

Основные положения логики-вероятностного метода

В основе подхода к решению задачи управления рисками лежит логики-вероятностный метод (ЛВМ), который состоит в сочетании логического и вероятностного подходов при оценке показателей надежности сложных технических, экономических, социальных и других систем [2].

В ЛВМ в качестве базовых используются понятия **опасного состояния системы** и **опасности** — способности системы переходить в опасное состояние. Описание опасного состояния системы начинается с составления **сценария опасного состояния** (ОС), который строится с использованием операций дизъюнкции и конъюнкции над **иницирующими условиями и событиями**.

В качестве инициирующих условий и событий выступают отказы одного или нескольких элементов системы. Каждому элементу системы ставится в соответствие **логическая переменная** x_k ($k = \overline{1, h}$) с двумя возможными состояниями (например, работоспособности/отказа, готовности/неготовности и т. п.) с заданными вероятностными параметрами этих состояний p_k и $q_k = 1 - p_k$.

Сценарий является основой для составления логической функции, или функции алгебры логики, описывающей опасное состояние системы.

Следующим шагом является преобразование функции алгебры логики к вероятностной функции, которая в дальнейшем используется для получения количественной оценки вероятности реализации опасного состояния [3].

Таким образом, с одной стороны, метод предоставляет механизм для формализации множества опасных состояний системы, а с другой стороны — теоретически обоснованный подход к количественной оценке риска системы.

Для системы, состоящей из различных ресурсов, ЛВМ используется в целях получения количественных оценок вероятностей реализации опасных состояний для каждого вида ресурсов. В свою очередь, каждый ресурс в ЛВМ также рассматривается как отдельная система.

Оценка интегрального риска системы на основе ЛВМ

Дано:

1. Ресурс с номером i , для которого выделены опасные состояния S_{ij} , $j = \overline{1, m}$, где m — число возможных состояний.

2. Структура ОС и вероятности инициирующих событий (угроз) x_k , $k = \overline{1, h}$.

Требуется найти:

1. Вероятности P_{ij} реализации опасных состояний i -го ресурса S_{ij} , $j = \overline{1, m}$.

2. Значимости $Z(x_k)$ каждого инициирующего условия или события x_k с учетом его вклада в реализацию опасного состояния S_{ij} ; k — номер инициирующего события или угрозы $k = \overline{1, h}$.

Алгоритм решения

Шаг 1. Составление сценария опасного состояния S_{ij} .

Шаг 2. Построение функции алгебры логики (ФАЛ) $f(x_1, \dots, x_h)$ с использованием операций конъюнкции и дизъюнкции на основе сценария опасного состояния S_{ij} .

Шаг 3. Построение вероятностной функции (ВФ) $P_{ij}\{f(x_1, \dots, x_h) = 1\}$ на основе ФАЛ.

Шаг 4. Расчет вероятности P_{ij} реализации ОС с помощью вероятностной функции.

Шаг 5. Расчет значимости $Z(x_k)$ каждой угрозы с учетом ее вклада в реализацию ОС.

Значимость элемента (угрозы) определяется на вероятностной модели как частная производная ВФ:

$$Z(x_k) = \frac{\partial P}{\partial P_k} = P\{\Delta_{xk} f(x_1, \dots, x_h) = 1\}.$$

Результатом данного этапа является рейтинг угроз по степени влияния на реализацию ОС.

Если задача решается для нескольких ОС, то под значимостью k -й угрозы для данного ОС понимается величина

$$\bar{Z}(x_k) = Z(x_k) \cdot s_k = \frac{\partial P}{\partial P_k} s_k = \frac{\partial r}{\partial P_k},$$

где s_k — ущерб в случае реализации ОС; r — ущерб для данного ОС.

Оценки вероятностей отказов ресурсов — либо статистически достоверные данные, либо экспертные данные.

Для подсчета ущерба используются известные подходы, такие как прямой счет и оценка пороговых значений риска [3].

Алгоритмы формирования перечня наиболее эффективных контрмер

Для дальнейшего рассмотрения отберем только те опасные состояния, для которых уровень риска $P_{ij} \geq P_0$, где P_0 — пороговое значение для уровня риска. Пороговое значение отражает критическую величину потерь от реализации опасного состояния.

Пронумеруем угрозы, отобранные для дальнейшего рассмотрения опасных состояний заново. Для этого введем новый индекс $v = \overline{1, V}$, учитывающий каждую из угроз для каждого опасного состояния каждого ресурса системы. При введении нового индекса учитывается тот факт, что одна и та же угроза может встречаться в нескольких

опасных состояниях. В этом случае под значимостью угрозы понимается сумма значимостей этой угрозы в различных опасных состояниях. Конечная цель — уменьшение уровня риска системы посредством реализации набора контрмер $\{g_t\}$, $t = \overline{1, T}$. Для контрмеры известна стоимость ее реализации s_t , также задана величина s_0 — бюджет, выделенный на превентивные мероприятия.

Далее представлены три варианта формирования перечня наиболее эффективных контрмер.

Вариант 1. Формирование перечня наиболее эффективных контрмер.

Представим информацию о том, влияет или не влияет данная контрмера на угрозу, в виде матрицы:

$$a_{tv} = \begin{pmatrix} a_{11} & \dots & a_{1V} \\ \dots & \dots & \dots \\ a_{T1} & \dots & a_{TV} \end{pmatrix},$$

где $a_{tv} \in \{0, 1\}$, $a_{tv} = 1$, если контрмера g_t влияет на угрозу x_v , и $a_{tv} = 0$, если контрмера g_t не влияет на угрозу x_v . Термин "влияет" означает "уменьшает вероятность реализации угрозы" или "устраняет угрозу".

Тогда задачу формирования перечня наиболее эффективных контрмер можно представить как задачу нахождения бинарного вектора $(b_1, b_2, \dots, b_T)$, такого что:

$$\sum_{t=1}^T \left(\sum_{v=1}^V z_v a_{tv} \right) b_t \rightarrow \max$$

при ограничениях $\sum_{t=1}^T s_t b_t \leq s_0$, $a_{tv} \in \{0, 1\}$, $b_t \in \{0, 1\}$,

$t = \overline{1, T}$, $v = \overline{1, V}$.

Алгоритм решения

Для решения задачи использовался направленный перебор на множестве бинарных векторов длины T . Далее для простоты приведен алгоритм полного перебора.

Задаем величину $\max = 0$, текущий вектор $b = (0, 0, \dots, 0, 1)$, результирующий вектор $r = (0, 0, \dots, 0, 1)$.

В цикле от 1 до 2^T выполняем следующие шаги.

Шаг 1. Проверяем бюджетное ограничение $\sum_{t=1}^T s_t b_t \leq s_0$ для текущего вектора b .

Шаг 2. Если ограничение выполнено, то вычисляем величину

$$C = \sum_{t=1}^T \left(\sum_{v=1}^V z_v a_{tv} \right) b_t,$$

иначе — переход к следующей итерации.

Шаг 3. Если $C > \max$, то $\max = C$, $r = b$.

Шаг 4. Формируем очередной вектор b .

Результатом работы алгоритма будет вектор r ,

на котором величина $C = \sum_{t=1}^T \left(\sum_{v=1}^V z_v a_{tv} \right) b_t$ достигает максимума.

Вариант 2. Формирование перечня наиболее эффективных контрмер.

Представим информацию о том, как влияет данная контрмера на угрозу, в виде матрицы

$$\Delta p_v^t = \begin{pmatrix} \Delta p_1^t & \dots & \Delta p_V^t \\ \dots & \dots & \dots \\ \Delta p_1^T & \dots & \Delta p_V^T \end{pmatrix},$$

где Δp_v^t — изменение вероятности реализации угрозы с номером v при попадании контрмеры с номером t в перечень контрмер для реализации, или 0, если контрмера с номером t не оказывает влияния на угрозу с номером v .

Тогда задачу формирования перечня наиболее эффективных контрмер можно представить как задачу нахождения бинарного вектора $(b_1, b_2, \dots, b_T)^T$, которому соответствует максимальное изменение интегрального риска системы $\Delta R \rightarrow \max$, согласно формулам (1)–(3), такого что:

$$\sum_{i=1}^n \left(\sum_{j=1}^m \Delta P_{ij} \cdot C_{ij} \right) \rightarrow \max$$

при ограничениях $\sum_{t=1}^T s_t b_t \leq s_0$, $b_t \in \{0, 1\}$, $r = \overline{1, T}$,
 $v = \overline{1, V}$.

Алгоритм решения

Для решения задачи также использовался направленный перебор на множестве бинарных векторов длины T . Ниже рассмотрен алгоритм полного перебора. Алгоритм состоит из 2^T итераций.

Задаем величину $\max = 0$, текущий вектор $b = (0, 0, \dots, 0, 1)$, результирующий вектор $r = (0, 0, \dots, 0, 1)$.

В цикле от 1 до 2^T выполняем следующие шаги.

Шаг 1. Проверяем бюджетное ограничение

$$\sum_{t=1}^T s_t b_t \leq s_0 \text{ для текущего вектора } b.$$

Шаг 2. Если ограничение выполнено, то вычисляем величину $\Delta R = \sum_{i=1}^n \left(\sum_{j=1}^m \Delta P_{ij} \cdot C_{ij} \right)$, иначе — переход к следующей итерации.

Шаг 3. Если $\Delta R > \max$, то $\max = \Delta R$, $r = b$.

Шаг 4. Формируем очередной вектор b .

Результатом работы алгоритма будет вектор r ,

на котором величина $\Delta R = \sum_{i=1}^n \left(\sum_{j=1}^m \Delta P_{ij} \cdot C_{ij} \right)$ достигает максимума.

Замечание. В варианте 2 для решения задачи необходимо знать Δp_v^t для всех контрмер, в том числе и для тех, которые не войдут в перечень контрмер для реализации, что требует дополнительных затрат.

Вариант 3. Формирование перечня наиболее эффективных контрмер.

Входные данные для этого алгоритма такие же, как для первого варианта.

Алгоритм решения

Шаг 1. Контрмеры группируем в f классов, критерием классификации контрмер является влияние на угрозу. Если представить эту информацию в виде бинарных векторов (0 — если контрмера не влияет на угрозу, 1 — если влияет), то элементы одного класса будут обладать одним и тем же набором единиц.

Шаг 2. В каждом классе выполняем ранжирование контрмер. В качестве критериев ранжирования принимаются стоимость, совокупная мера влияния на угрозы.

Шаг 3. Выделенные классы контрмер упорядочиваем по эффективности Q :

	x_1	x_2	$\dots$	x_v	Q
G^1	1	0	$\dots$	0	l_1
G^2	0	1	$\dots$	0	l_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
G^f	1	0	$\dots$	1	l_f

где $G^1, G^2, \dots, G^f$ — классы контрмер, Q — показатель эффективности класса контрмер, $l_i = \sum_v z_{iv}$, f — число классов контрмер.

Шаг 4. Составляем вектор вида $(b_i^1, b_j^2, \dots, b_k^w)^T$ из лучших контрмер каждого класса (где, например, b_j^2 означает, что j -я контрмера после упорядочивания во втором классе включена в набор рассматриваемых контрмер). Проверяем условие

$$\sum_{i=1}^f S_p^i \leq S_0, \quad (4)$$

где p — номер контрмеры, выбранной из i -го класса.

Если оно выполняется, то текущий вектор является решением. Если условие не выполняется, то переходим к шагу 5.



Рис. 1. Функциональная модель задачи управления (снижения) информационных рисков

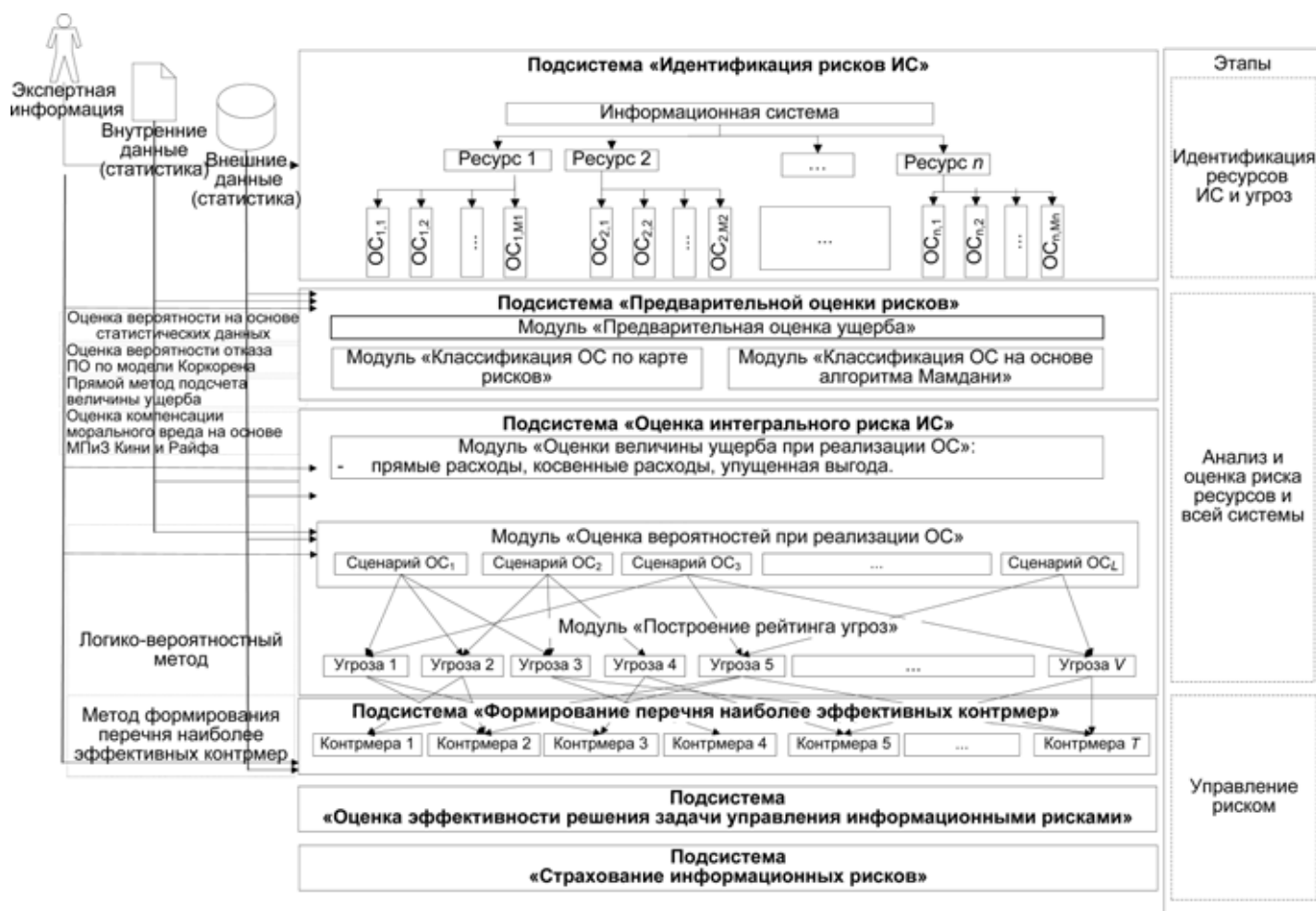


Рис. 2. Структура программного комплекса "СУИР ЛВМ"

Шаг 5. Уменьшаем суммарную эффективность текущего вектора путем замены в последнем классе выбранной контрмеры на контрмеру меньшего ранга, начиная с последнего класса контрмер.

Увеличиваем число классов для одновременной замены в них контрмер, затем выбираем контрмеры более низкого ранга.

Оценка уровня риска системы с учетом контрмер и оценка эффективности решения задачи управления рисками

После получения вектора $(b_1, b_2, \dots, b_T)^T$ можно рассчитать уровень риска по каждому опасному состоянию, ресурсу и всей системе.

При расчете уровня риска опасных состояний x_k^{ij} сначала нужно определить вероятности реализации угроз p_k^{ij} с учетом выбранных для реализации контрмер:

$$(p_k^{ij})' = \begin{cases} p_k^{ij}, & \text{если } P_{ij} \leq P_0; \\ 0, & \text{если } \exists G_p, \text{ которое устраняет угрозу;} \\ p_k^{ij} - \Delta p_k^{ij}, & \text{если } \exists G_p, \text{ которое} \\ & \text{уменьшает вероятность реализации} \\ & \text{угрозы на } \Delta p_k^{ij}. \end{cases}$$

В остальном алгоритм расчета уровня риска системы не меняется.

Пусть уровень риска системы с учетом выбранных для реализации контрмер равен $\bar{R}$, тогда эффективность Ef выбранных для реализации контрмер примет вид:

$$Ef = \frac{R - \bar{R}}{R} 100 \ \%.$$

Авторы реализовали предложенный подход риск-анализа для управления информационными рисками. Был разработан и внедрен программный продукт "Система управления информационными

рисками на основе ЛВМ" ("СУИР ЛВМ"). На рис. 1 представлен фрагмент функциональной модели управления (снижения) информационными рисками системы, построенной в программе VRwin.

Разработанная "СУИР ЛВМ" является представителем класса систем анализа защищенности и управления риском. Данная программа позволяет моделировать риски безопасности средних и крупных информационных систем организаций с высокой степенью адекватности. Система позволяет рационально распределить время на оценку ИР: игнорировать малозначимые риски, быстро оценивать средние риски и с высокой степенью детализации и точности оценивать наиболее значимые риски. На рис. 2 представлены основные подсистемы "СУИР ЛВМ".

Пользователями системы являются специалисты по информационной безопасности, системные администраторы, аудиторы и андеррайтеры страховых компаний, занимающихся страхованием информационных рисков.

Заключение

Предложенные алгоритмы позволяют наряду с другими методами риск-анализа снижать потенциальный ущерб за счет формирования оптимального состава превентивных мер при заданном бюджете. Эффективность решения обусловлена применением ЛВМ на всех этапах управления и оценки рисков с различной степенью детализации опасных состояний системы.

Список литературы

1. Cronin M. J. 2000. Privacy and the Internet [Электронный ресурс]. Режим доступа: <http://www.hoover.stanford.edu/publications/digest/003/cronin.html>.
2. Рябинин И. А. Надежность и безопасность структурно-сложных систем. СПб.: Политехника, 2000. 248 с.
3. Чернова Г. В., Кудрявцев А. А. Управление риском: Учебн. пособие. М.: ТК Велби, Изд-во Проспект, 2006. 160 с.

С. В. Жернаков, д-р. техн. наук, проф.,
Уфимский государственный авиационный
технический университет,
e-mail: zhsviit@mail.ru

Восстановление потерянной информации с датчиков ГТД в условиях бортовой реализации с использованием технологии нейронных сетей

Рассматривается один из подходов к восстановлению потерянной информации с датчиков в условиях бортовой реализации на основе нейросетевой парадигмы. Приводится методика решения таких задач в нейросетевом базисе.

Ключевые слова: отказоустойчивость, контроль, диагностика, нейронные сети

Особенности бортовой реализации нейросетевых алгоритмов контроля и диагностики ГТД

Анализ работ в области построения бортовых систем контроля и диагностики (БСКД) современных авиационных двигателей как в России, так и за рубежом свидетельствует о том, что при их разработке наметилась устойчивая тенденция к интеллектуализации и переход от пассивного контроля к активному, позволяющему не только локализовать отказы, но и по возможности устранить их [1–3]. Таким образом, БСКД постепенно выделяются в самостоятельные системы, обладающие высокой сложностью и многопроцессорностью, способные решать комплексные задачи с высоким качеством и эффективностью в темпе реального времени. Следует отметить и то обстоятельство, что современные БСКД взаимодействуют с системами управления ГТД, позволяя последним, плавно и своевременно воздействуя на исполнительные механизмы, с одной стороны, повышать качество управления авиационным двигателем и его подсистемами, а с другой — повышать надежность ГТД в процессе его эксплуатации. Однако наличие факторов неопределенности (НЕ-факторов), а также необходимость принятия оперативных решений в реальном времени

создают дополнительные проблемы в процессе контроля и диагностики в рамках БСКД.

Анализ современных БСКД показывает, что спектр задач, решаемых ими, непрерывно растет. Это объясняется постоянным совершенствованием самого авиационного двигателя и, как следствие, выполняемых им функций, а также соответствующим увеличением числа контролируемых и диагностируемых параметров на борту самолета. Так, БСКД для ТВВД [1,3] решает широкий спектр задач, связанных с оценкой остаточного ресурса ГТД и контролем его параметров. В процессе решения задачи выработки ресурса ГТД решаются задачи прогноза (краткосрочного) текущего состояния двигателя и тренд-анализа его параметров. В БСКД ТВВД решаются задачи контроля большого числа параметров двигателя:

- температуры и давления газов за и перед турбиной;
- частоты вращения ротора компрессора низкого и высокого давления;
- работоспособности системы смазки и суфлирования ГТД по предельным значениям параметров маслосистемы двигателя;
- расхода масла по изменению уровня в маслобаке ГТД;
- работоспособности топливной системы ГТД;
- уровня вибраций;
- температуры в полостях двигателя;
- температуры подшипника;
- массового расхода топлива;
- температуры и давления воздуха за компрессором;
- технического состояния проточной части двигателя по его термогазодинамическим параметрам.

Вместе с тем, при большом многообразии решаемых задач существуют определенные сложности:

- ограниченные вычислительные ресурсы (объем оперативной памяти, быстродействие, точность полученных результатов);
- трудности формализации классических алгоритмов контроля и диагностики и, как следствие, их практической реализации;
- необходимость реализации перечисленных выше алгоритмов на языке низкого уровня (ассемблер);
- трудности восстановления информации при отказе датчиков.

В этих условиях использование нейросетевых технологий весьма перспективно. Ниже рассматриваются особенности применения нейросетевых технологий в составе БСКД на примере решения последней из перечисленных выше задач, связанной с использованием имеющейся аналитической избыточности при измерении параметров ГТД для восстановления сигналов при отказах датчиков.

Восстановление информации с использованием нейросетевой модели ГТД

Одной из ключевых задач, решаемых на борту летательного аппарата, является выявление параметрических отказов датчиков (деградации их характеристик). В этом случае может, как правило, использоваться метод мажоритарного контроля (рис. 1) [4].

Суть метода мажоритарного контроля заключается в том, что если среди датчиков D_1 или D_2 (измеряющих, например, давление воздуха за компрессором P_2^*) (рис. 1) один из них имеет характеристику, отличную от эталонной, то введение дополнительного канала измерения в виде математической модели (ММ) позволяет вычислять среднее значение данного измерения (медиану) и тем самым восстанавливать потерянную с

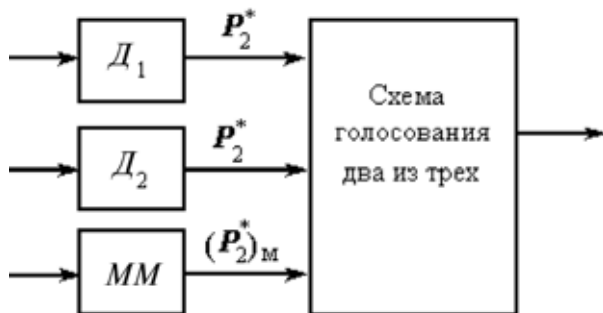


Рис. 1. Реализация мажоритарного контроля

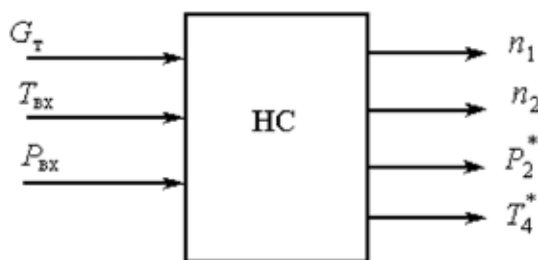


Рис. 2. Нейросетевая модель ГТД: G_t — расход топлива, подаваемого в камеру сгорания (кг/с); $T_{вх}$ — температура на входе в двигатель (К); $P_{вх}$ — давление на входе в двигатель (Па); n_1 — частота вращения ротора компрессора низкого давления (%); n_2 — частота вращения ротора компрессора высокого давления (%); P_2^* — давление воздуха за компрессором (КПа); T_4^* — температура газов за турбиной (К)

датчика информацию. В данной работе предлагается использовать в качестве ММ ГТД его нейросетевую модель, полученную в процессе решения задачи идентификации (рис. 2).

Решение задачи восстановления информации с использованием нейросетевой модели ГТД декомпозируется на две подзадачи:

- задача идентификации ММ (подробно рассматривается в № 3 журнала "Информационные технологии" за 2010 г.);
- задача восстановления информации.

Рассмотрим подробно решение задачи восстановления информации в условиях бортовой реализации.

Восстановление информации с использованием технологии нейронных сетей

Ниже подробно остановимся на решении задачи восстановления информации с использованием нейронной сети (НС). В качестве НС был выбран персептрон, имеющий структуру 3 — 6 — 4. В качестве алгоритма обучения использован метод сопряженных градиентов.

Проведем анализ вычислительных затрат нейросетевого алгоритма реализации ММ ГТД на базе микропроцессорного контроллера STG Thompson ST10F269 и нейрочипа Л1879ВМ1 (NM6403). Их технические характеристики приведены в табл. 1.

Из табл. 1 видно, что микроконтроллеры имеют одинаковые тактовые частоты, но операции умножения и деления выполняются ими за различные интервалы времени. Это связано с разными архитектурами микроконтроллеров. Так, Л1879ВМ1 построен на основе RISC-архитектуры, что подразумевает выполнение любой команды за один такт. Вместе с тем, объем его ОЗУ намного меньше, чем у ST10F269.

Расчет времени работы с памятью. Время обращения к памяти при чтении/записи $T_{обр}$ рассчитывается следующим образом:

$$T_{обр} = T_a + T_d + T_y, \quad (1)$$

где $T_{обр}$ — время обращения к памяти; T_a — время готовности шины адреса; T_d — время готовности шины данных; T_y — время готовности шины управления.

Для Л1879ВМ1 число обращений к памяти будет в два раза меньше по сравнению с ST10F269, так как он работает с 64-разрядной памятью, а ST10F269 — с 32-разрядной памятью. Таким образом, для ST10F269 при работе с памятью необходимы три такта: запись из регистра в память; запись данных в регистр данных; установка управляющего слова.

Согласно (1) имеем:

$$T_{обр} = 25 + 25 + 25 = 75 \text{ нс.}$$

Таблица 1

Технические характеристики микроконтроллеров

Параметр	ST10F269	Л1879ВМ1 (NM6403)
Тактовая частота	40 МГц	40 МГц
Максимальное время выполнения команды сложения	25 нс	25 нс
Максимальное время выполнения команды умножения	250 нс	25 нс
Максимальное время выполнения команды деления	500 нс	25 нс
Разрядность (рег/данных)	16/32	32/64
Объем внутреннего ОЗУ	12 Кбайт	256 байт
Объем внутреннего ПЗУ	256 Кбайт	—

Расчет объема данных для хранения нейронной сети. Размер памяти, необходимый для НС в ОЗУ, вычисляется следующим образом:

$$V = \left(\sum_{i=1}^3 N_i + N_c + \sum_{i=1}^2 N_i N_{i+1} \right) U, \quad (2)$$

где N_i — число нейронов в i -м слое НС; N_c — число слоев НС; $\sum_{i=1}^2 N_i N_{i+1}$ — число весовых коэффициентов НС; U — число разрядов при записи смещения нейрона (4 байт).

Тогда для НС, имеющей структуру 3 — 6 — 4, согласно (2) получим:

$$V = ((3 + 6 + 4) + 3 + (3 \cdot 6 + 6 \cdot 4))4 = 232 \text{ байт.}$$

Расчет числа выполняемых операций. В процессе работы НС выполняется ряд математических (сложение, умножение, деление) и логических (сравнение) операций.

При этом вычисляется взвешенная сумма входов нейрона, к которой добавляется смещение. Полученный результат подается на функцию активации нейрона. В качестве последней в данном случае выбирается рациональная сигмоида $z = f(x)$, вычисляемая по формуле [5, 6]

$$z = \frac{x}{|x| + c}, \quad (3)$$

где c — крутизна сигмоидной функции, при расчетах принималось $c = 1$.

При этом, если на нейропроцессоре Л1879ВМ1 функции активации нейронов реализуются аппаратно, то на микроконтроллере ST10F269 они выполняются программно. Для рациональной сигмоиды (3) получим две операции — сложение и деление. Всего в процессе работы НС будет выполнено следующее число операций:

$$OS = N_1 + \sum_{i=1}^2 N_i N_{i+1} + \sum_{i=1}^3 N_i; \quad (4)$$

$$OU = N_1 + \sum_{i=1}^2 N_i N_{i+1} + \sum_{i=1}^3 N_i; \quad (5)$$

$$Tf = \sum_{i=2}^3 N_i, \quad (6)$$

где OS — число операций сложения; OU — число операций умножения; Tf — число функций активации; N_i — число нейронов в i -м слое НС.

Тогда для рассмотренной выше структуры НС согласно (4)–(6), имеем:

$$OS = 3 + (3 \cdot 6 + 6 \cdot 4) + (3 + 6 + 4) = 58;$$

$$OU = 3 + (3 \cdot 6 + 6 \cdot 4) + (3 + 6 + 4) = 58;$$

$$Tf = 10.$$

Считая, что все операции в микроконтроллере Л1879ВМ1 выполняются за один такт, включая функцию активации нейрона, а также учитывая разрядность, скорость обращения к памяти и параллельность выполнения алгоритма, имеем общее число операций:

$$O_{\text{Л1879ВМ1}} = 58/2 + 58/2 + 3/2 + 6/2 + 4/2 \approx 64.$$

Расчет времени реализации НС. Для микроконтроллера ST10F269 время выполнения всех операций (с учетом времени выборки данных из памяти) можно вычислить как

$$T_{\text{ST10F269}} = Od \cdot t_d + OS \cdot t_p + OU \cdot t_m + Tf \cdot t_p + V_{nn} \cdot t_{mem}, \quad (7)$$

где t_d — время выполнения команды деления; t_p — время выполнения команды сложения; t_m — время выполнения команды умножения; t_{mem} — время выполнения выборки данных из памяти.

Тогда, согласно (7), для ST10F269 получим:

$$T_{\text{ST10F269}} = 10 \cdot 500 + 58 \cdot 25 + 58 \cdot 250 + 10 \cdot 25 + 160 \cdot 25 = 25200 = 25,2 \text{ (мкс)}.$$

Учитывая разную разрядность процессора (16 разрядов) и операндов (32 разряда), время вычисления увеличится в два раза:

$$T_{\text{ST10F269}} = 50400 = 50,4 \text{ (мкс)}.$$

Для нейропроцессора Л1879ВМ1 это время составит:

$$T_{\text{Л1879ВМ1}} = 64 \cdot 25 + 80 \cdot 25 = 3600 = 3,6 \text{ мкс}.$$

Отсюда видно, что специализированный процессор Л1879ВМ1 выполняет реализацию НС в 14 раз быстрее вычислителя на базе STG Thompson ST10F269.

Результаты сравнительного анализа нейросетевого метода с аналитической (кусочно-линейной) моделью ГТД, а также полиномиальной регрессионной моделью, реализованной на основе метода наименьших квадратов (полином девятого порядка), приведены в табл. 2.

Таблица 2

**Сравнительный анализ различных способов
решения задачи идентификации ГТД**

Модели	Время вычисления, мкс	Объем занимаемой памяти, байт	Погрешность вычисления, %			
			n_1	n_2	T_4^*	P_2^*
Кусочно-линейная модель	46,3	3192	3,71	1,57	2,88	6,32
Полиномиальная модель	51,4	486	1,14	0,97	1,56	2,61
Нейросетевая модель	3,6	394	0,75	0,68	0,81	0,76

Данные, приведенные в табл. 2, получены путем сравнения реальных данных летных испытаний с данными, вычисленными по указанным выше моделям двигателя, и соответствуют изменению режимов его работы от "малого газа" до режима "максимал".

Анализ табл. 2 показывает, что нейросетевая модель ГТД эффективнее решает задачу идентификации многорежимного ГТД: максимальная погрешность идентификации не превысила 0,81 %, что в 7,8 раза меньше максимальной погрешности кусочно-линейной модели и в 3 раза меньше максимальной погрешности полиномиальной модели ГТД. Таким образом, полученная нейросетевая модель может эффективно использоваться в составе современных бортовых систем контроля и диагностики для восстановления информации при отказах датчиков.

**Восстановление информации
при отказах датчиков с использованием
автоассоциативной нейронной сети**

Другим альтернативным подходом к восстановлению информации является использование для этих целей автоассоциативной нейронной сети (АНС), обладающей свойством сжатия и последующего восстановления информации.

Известны работы [6—8], в которых показана возможность применения АНС для восстановления измеренной информации о термогазодинамических параметрах ГТД. Вместе с тем, в этих работах отсутствуют методики их построения и рекомендации по выбору структуры и алгоритмов обучения АНС.

Ниже исследуются особенности формализации и решения задачи восстановления информации при отказах датчиков ГТД с помощью АНС и бортовой реализации соответствующих алгоритмов.

Будем рассматривать следующую задачу: требуется построить АНС, обеспечивающую контроль и восстановление информации (в случае возникновения отказов) о техническом состоянии ГТД при наличии пяти штатных датчиков термогазодинамических параметров G_T , n_1 , n_2 , P_2^* , T_4^* , где G_T — расход топлива (кг/с); n_1 — частота вращения ротора компрессора низкого давления; n_2 — частота вращения ротора компрессора высокого давления; P_2^* — давление воздуха за компрессором (КПа); T_2^* — температура воздуха за компрессором (К); T_4^* — температура газов за турбиной (К).

На рис. 3 приведена архитектура АНС, состоящей из трех слоев. АНС осуществляет зеркальное отображение вектора входных данных на самого себя. Сжатие информации происходит в скрытом слое НС, называемом "горлом", восстановление информации осуществляется в выходном слое. Принципиальная возможность использования сжимающего отображения составляет основу метода главных компонент.

Ключевым моментом в выборе структуры АНС является определение оптимального числа нейронов в горле. На рис. 4 приведен график зависимости ошибки восстановления информации сетью при отказе одного из датчиков (в данном случае — датчика частоты вращения n_1) от числа нейронов в горле АНС.

По оси абсцисс на рис. 4 отложен относительный размер горла — отношение числа нейронов в горле к числу входных/выходных нейронов (пять нейронов). Как видно на рисунке, оптимальный размер горла в данном случае равен 0,8, т. е. число нейронов в горле (скрытом слое АНС) должно равняться четырем.

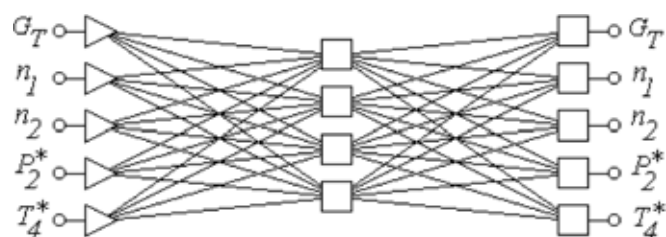


Рис. 3. Архитектура АНС

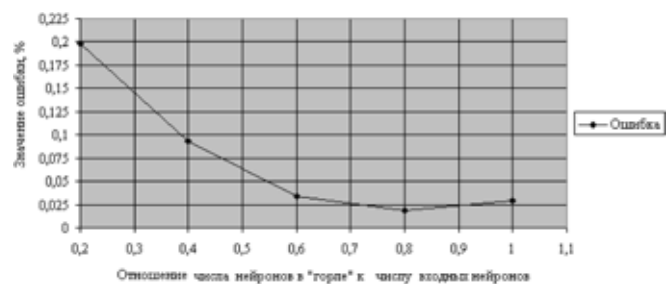


Рис. 4. Выбор оптимальной структуры АНС



Рис. 5. Функциональная схема восстановления информации на основе АНС

Данные для обучения АНС были получены в процессе летных испытаний авиационного двигателя в широком диапазоне изменения режимов его работы (начиная от малого газа до форсажного режима).

Наряду с вопросом выбора структуры АНС важным этапом является этап восстановления информации по результатам функционирования этой нейронной сети.

В данной работе используется подход, предложенный в [6] и изображенный в виде функциональной схемы на рис. 5.

Сигналы с датчиков одновременно поступают в стек и систему допускового контроля, осуществляющую контроль измеренных значений с датчиков. Если сигналы с датчиков находятся в поле допуска, то они беспрепятственно передаются далее в систему управления, которая осуществляет управление исполнительными механизмами (ИМ), воздействующими на ГТД. В случае, если какой-либо из сигналов датчиков выходит за поле допуска, то система допускового контроля извлекает последнее значение этого сигнала из стека и передает его на соответствующий вход АНС, которая восстанавливает информацию и далее передает ее в САУ авиационным двигателем.

Результаты восстановления информации АНС в случае отказа датчика P_2^* приведены в табл. 3.

При отказе других датчиков погрешность восстановления не превышает значений, приведенных в табл. 3.

Анализ полученных результатов показывает высокую эффективность применения АНС в процессе восстановления информации.

Результаты восстановления информации в случае двойного (кратных) отказов датчиков n_1 и n_2 приведены в табл. 4.

Анализ результатов, приведенных в табл. 4, показывает, что несмотря на то, что точность восстановления информации в случае двойных отказов датчиков ухудшается, применение АНС обеспечивает удовлетворительное качество восстановления информации, т. е. сохранение работоспособности измерительных каналов САУ ГТД.

Проведем анализ требуемых вычислительных затрат, необходимых для реализации нейросетевой модели ГТД на базе микропроцессорного контроллера STG Thompson ST10F269 и нейрочипа Л1879ВМ1 (NM6403) (см. табл. 1).

Время обращения к памяти (чтение/запись) рассчитывается аналогично расчету по формуле (1).

Размер памяти ОЗУ, необходимый для вычислений с помощью НС, вычисляется по формуле (2):

$$V = ((5 + 4 + 5) + 3 + (5 \cdot 4 + 4 \cdot 5)) \cdot 4 = 228 \text{ байт.}$$

В процессе вычисления выходов НС будет выполнено следующее число операций, вычисляемых на основании формул (4)–(6):

$$\begin{aligned} OS &= 3 + (5 \cdot 4 + 4 \cdot 5) + (5 + 4 + 5) = 57; \\ OU &= 3 + (5 \cdot 4 + 4 \cdot 5) + (5 + 4 + 5) = 57; \\ Tf &= 9. \end{aligned}$$

Считая, что разрядность операндов равна 32, а разрядность процессора Л1879ВМ1 равна 64, а также с учетом того, что все операции им выполняются за один такт, включая вычисление функции активации нейрона (рациональная сигмоида), для выполнения алгоритма необходимо число выполняемых операций поделить на два:

$$O_{\text{Л1879ВМ1}} = 57/2 + 57/2 + 5/2 + 4/2 + 5/2 = 62.$$

Для микроконтроллера ST10F269 время выполнения всех операций на основании (7) составит:

$$T_{\text{ST10F269}} = 9 \cdot 500 + 57 \cdot 25 + 57 \cdot 250 + 9 \cdot 25 + 160 \cdot 25 = 24400 \text{ нс} = 24,4 \text{ мкс.}$$

Учитывая разрядность процессора (16 разрядов) и разрядность операндов (32 разряда), время выполнения увеличится в два раза:

$$T_{\text{ST10F269}} = 48,8 \text{ мкс.}$$

Таблица 3

Качество восстановления информации АНС при отказе одного из датчиков (P_2^*)

Структура АНС	Погрешность восстановления параметров ГТД, %				
	G_T	n_1	n_2	P_2^*	T_4^*
5—4—5	0,33	0,31	0,24	0,45	0,38

Таблица 4

Качество восстановления информации АНС при отказе двух датчиков (n_1 и n_2)

Структура АНС	Погрешность восстановления параметров ГТД, %				
	G_T	n_1	n_2	P_2^*	T_4^*
5—4—5	0,56	0,45	0,44	0,52	0,58

Для Л1879ВМ1 это время составит:

$$T_{\text{Л1879ВМ1}} = 58 \cdot 25 + 80 \cdot 25 = \\ = 3450 \text{ (нс)} = 3,4 \text{ мкс.}$$

Отсюда следует, что специализированный микроконтроллер выполняет реализацию НС в 14 раз быстрее бортового вычислителя на базе STG Thompson ST10F269.

Таким образом, использование АНС в БСКД для восстановления информации позволяет обеспечить отказоустойчивость измерительных каналов систем управления ГТД.

Основным преимуществом применения нейронных сетей в рамках БСКД является возможность обучения и дообучения в режиме реального времени, с учетом индивидуальных характеристик конкретного авиационного двигателя.

Выводы

- На основании проведенных исследований сформулированы требования к нейросетевой системе контроля и диагностики технического состояния ГТД, позволяющей использовать, наряду с количественными математическими моделями двигателя, качественную и экспериментальную информацию, полученную в ходе стендовых и летных испытаний ГТД.
- Рассмотрены особенности реализации нейросетевых алгоритмов восстановления информации в случае отказа штатных датчиков в составе бортовой системы контроля и диагностики двигателя (БСКД). На примере БСКД ТВВД показано, что:
 - реализация нейросетевой многорежимной модели ГТД обеспечивает уменьшение погрешности вычисления термогазодинамических параметров ГТД в $2,3 \div 7,8$ раза по сравнению с кусочно-линейной моделью ГТД и в $1,4 \div 3,4$ раза по сравнению с полиномиальной моделью 9-го порядка; при этом за счет сокращения числа операций умножения обеспечивается выигрыш по сравнению с кусочно-линейной и полино-

миальной моделями при использовании вычислителя STG Thompson ST10F269 соответственно: по времени вычислений в 12,8 и в 14,2 раза; по требуемому объему памяти в 8,1 и 1,2 раза (при использовании нейропроцессора Л1879ВМ1 — время вычисления НС дополнительно уменьшается в 2 раза);

- восстановление информации при отказе датчиков с помощью АНС обеспечивает погрешность восстановления данных не более 0,45 % в случае одиночных отказов и не более 0,58 % в случае двойных отказов; при этом время вычисления одного цикла восстановления данных при использовании пяти штатных датчиков составляет 48,8 мкс для вычислителя STG Thompson ST10F269 и 3,4 мкс для специализированного нейропроцессора Л1879ВМ1, что удовлетворяет требованиям бортовой реализации в составе перспективной БСКД.

Результаты научных исследований поддержаны грантами РФФИ: № 06-08-00446-а, 08-08-00774-а.

Список литературы

1. Васильев С. Н. и др. Интеллектуальное управление динамическими системами. М.: Физматлит, 2000. 352 с.
2. Котенко П. С. Системы технического обслуживания и автоматизированного контроля летательных аппаратов: Учеб. пособие. Уфа: УГАТУ, 2000. 212 с.
3. Федун Б. Е. Проблемы разработки бортовых оперативно-советующих экспертных систем для антропоцентрических объектов // Изв. РАН. Теория и системы управления. 1996. № 5. С. 147—159.
4. Васильев В. И. и др. Автоматический контроль и диагностика систем управления силовыми установками летательных аппаратов. М.: Машиностроение, 1989. 240 с.
5. Головкин В. А. Нейронные сети: Обучение, организация и применение. М.: Радиотехника, 2001. 256 с.
6. Валеев С. С., Васильев В. И., Ильясов Б. Г. Отказоустойчивые системы управления сложными динамическими объектами с использованием искусственных нейронных сетей // Нейрокомпьютеры: разработка, применение. 2000. № 1. С. 32—35.
7. Мерил У. К. Обнаружение отказов датчиков ГТД за счет аналитической избыточности // Аэрокосмическая техника. 2004. № 6. С. 28—41.
8. Modern Problems of Design of Aeroengine Control Systems / Eds: Sun Jianguo, V. I. Vasilyev, B. G. Ilyasov, Beijing, China, 2004.

О. И. Глушенко, канд. эконом. наук,
ст. преподаватель,

Стерлитамакский институт
физической культуры,
e-mail: guzsa@ufamts.ru

С. А. Горбатов, д-р техн. наук, проф.,
Всероссийский заочный

финансово-экономический институт, г. Уфа,

Нейросетевая модель поддержки принятия решений в процессах муниципального бюджетирования

Разработана экономическая концепция эффективного краткосрочного планирования расходной части бюджета муниципальных образований и принятия соответствующих управленческих решений, которая позволяет стимулировать объекты нижнего уровня муниципальных учреждений к повышению эффективности своей работы. Применены методы системного анализа и нейросетевых инструментариев для моделирования такой сложной проблемы, как муниципальное бюджетирование.

Ключевые слова: муниципальное бюджетирование, фонд материального развития, нейросетевая модель, субвенции, бюджетные отношения, муниципальные образования, муниципальные учреждения, бюджетные услуги

Статья посвящена разработке научно-обоснованного подхода к распределению фонда материального развития (ФМР) между муниципальными учреждениями (МУ) для субвенций, направляемых муниципальным образованиям (МО) из бюджета РФ, на базе динамической нейросетевой модели.

Актуальность статьи имеет два аспекта — экономический и информационно-технологический. Экономический аспект заключается в том, что существующая нормативная методика распределения субвенций [1] между МО незамкнута "снизу", т. е. на уровне МУ. Информационно-технологический аспект обусловлен сложными условиями моделирования процесса бюджетирования муниципального уровня [2] при использовании нейросетевых инструментариев с сильным зашумлением входных данных, отягченным дефицитом наблюдений. Поэтому требуется привнесение в технологию нейросетевого моделирования ряда новых идей для парирования негативного воздействия сложных условий моделирования на качество модели. Эти идеи будут подробно раскрыты ниже.

1. Экономическая концепция планирования ФМР муниципальных учреждений

Предлагаемую экономическую концепцию можно кратко сформулировать так: *на нижнем уровне III системы (рис. 1), т. е. на уровне МУ, целесообразно реализовать принцип прозрачного объективного распределения ФМР выделенного данного МО между его бюджетополучателями МУ, пропорционально обобщенному показателю Φ эффективности работы этих МУ в течение некоторого ретроспективного периода (3...5 лет) с прогнозом показателя Φ на планируемый период.* Показатель Φ должен быть адаптирован для каждого вида бюджетных услуг населению. Например, для общеобразовательных средних учебных заведений таким показателем может служить качество учебного процесса, для МУ здравоохранения — это продолжительность жизни людей и качество медицинского обслуживания и т. д. Авторы считают, что предлагаемая концепция естественным образом вытекает из методических рекомендаций [1], "замыкая" их снизу на уровне МУ. Система, показанная на рис. 1, позволяет взвешенно и объективно учитывать как интересы Российской Федерации по решению национальных проектов в области образования, здравоохранения, интересы МО, а также интересы МУ, непосредственно обслуживающих население. Предлагаемая трехуровневая система распределения субвенций учитывает особенности внешней экономической среды на территориях, где расположены МО, финансово-экономическое состояние каждого МУ и показатели эффективности его работы.



Рис. 1. Структурная схема трехуровневой системы распределения субвенций

Сделаем дополнительные комментарии к схеме рис. 1. Верхний уровень I иерархической схемы реализует принцип распределения всего объема выделяемых финансовых средств (субвенций) между МО. Этот принцип можно назвать "*принципом равномерной нормативной подушевой обеспеченности*" населения МО в бюджетных услугах определенного вида, например общеобразовательных.

На уровне I используется простая детерминированная модель вида [1], в основу которой положены численность постоянного населения МО и нормативы, определенные законодательством РФ и субъектов РФ, по обеспечению каждого гражданина бюджетными услугами определенного вида.

Реализация принципа "равномерности" реализует гармоническое развитие территорий МО — "подтягивание" отстающих территорий МО к городскому уровню обеспеченности бюджетными услугами — и тем самым создает предпосылки для решения ряда национальных проектов.

Средний уровень II реализует принцип распределения ФМР между МО, т. е. той части субвенций, которая направлена на выравнивание фактически сложившегося уровня обеспеченности МО бюджетными услугами. Здесь учитывается предыстория процесса развития муниципальных инфраструктур. Данный принцип можно назвать "*принципом обеспечения равномерности темпов развития МО*", поскольку выделяемый МО ФМР пропорционален разности между средним по субъекту РФ показателем уровня потребности в определенных объектах социальной и инженерной инфраструктуры в МО в расчете на одного жителя и фактическим показателем этой потребности в данном МО.

Модель [1] распределения на уровне II также является детерминированной с использованием экспертных поправочных коэффициентов, учитывающих структуру населения и территориальные различия в стоимости создания единицы соответствующих объектов социальной и инженерной инфраструктуры.

Сформулированный выше принцип распределения ФМР на уровне III можно назвать "*соревновательным принципом*", поскольку в лимитируемых пределах верхнего и нижнего уровней ФМР (устанавливается постановлениями местных законодательных органов) больше получает то МУ, у которого прогнозный обобщенный показатель эффективности Φ выше на планируемый период.

На нижнем уровне III модель наиболее сложная из всех уровней — это гибридная оптимизационно-стохастическая модель, включающая в себя три части:

- нейросетевую модель (НСМ), которая восстанавливает скрытую в данных динамическую,

нелинейную многофакторную функциональную зависимость обобщенного показателя эффективности Φ от ряда первичных входных факторов $X_j, j = \overline{1, n}$, ответственных за финансово-экономическое состояние МУ;

- оптимизационную модель для задачи линейного программирования (ЗЛП), в которой количественно находит отражение "*соревновательный принцип*";
- прогнозную модель временных рядов (ВР), по которой определяются в точке прогноза все динамические входные факторы $X_j(t)$, введенные в НСМ.

2. Постановка задачи и предобработка данных

Изложение основных идей предлагаемого подхода к построению модели уровня III схемы (см. рис. 1) сделаем для конкретности на примере распределения ФМР между средними общеобразовательными учреждениями, однако полученные результаты носят достаточно общий характер, применительно к процессам муниципального бюджетирования.

Построим НСМ обобщенного показателя эффективности $\Phi(X(t), t)$ для достаточно однородного кластера $\{МУ_j\}_{j=1}^{j=n}$ как многомерную функцию вектора первичных факторов $X = (X_1, X_2, \dots, X_k, \dots, n)$, изменяющихся во времени. Здесь j — номер МУ в кластере. Термин "достаточно однородный кластер" понимается в смысле процедуры образования оптимального числа кластеров из работ [2, 3] при предобработке данных.

Введение в модель $\{f_j\}$ связывает интересы верхнего I и нижнего III уровней бюджетирования и стимулирует развитие МУ: чем выше эффективный показатель $\{f_j\}$, тем большая доля средств будет выделена j -му МУ по ФМР.

Формально НСМ можно записать в виде:

$$\Phi(X(t), t) = F(X(t), t, W) + l, \quad (1)$$

где $\Phi(X(t), t)$ — восстанавливаемая функциональная зависимость; $F(X(t), t, W)$ — нейросетевая аппроксимация условного математического ожидания $M|\Phi|_{X, t}$ при условии, что вектор X и время t приняли фиксированное значение; l — случайные остатки; W — матрица параметров НСМ — синаптических весов и смещение состояния нейронов [4].

Спецификация переменных НСМ проводилась на основе итерационной процедуры экспертных оценок с "дообучением экспертов" из работы [5]. Суть этой процедуры состоит в том, что строится ряд промежуточных нейросетевых субмоделей (НССМ) с разными экспертно выбранными наборами переменных $\{Y, X\}$. На каждом шаге итерации экспертам предъявляются результаты моделирования по НССМ. При этом степень по-

нимания моделируемого процесса экспертным комитетом увеличивается и, соответственно, модифицируется состав переменных. Качество НССМ на каждом шаге итерации отслеживается по векторному критерию, включающему в себя ошибку обобщения сети и коэффициент детерминации $r_{\hat{Y}, \hat{Y}}^2$, где $\hat{Y}$, $\hat{Y}$ — соответственно экспериментальное и расчетное по формуле (1) значение моделируемого показателя.

Этот метод позволяет в некоторой мере упорядочить (частично формализовать) внесение в модель априорной информации о моделируемом объекте. Достижимый эффект от процедуры дообучения экспертов в отношении выбора моделируемого показателя — это обострение их интуиции, отображаемой в назначении весовых множителей при образовании агрегированных переменных [см. ниже формулы (2) и (3)], а также в выборе состава независимых переменных $\{X_k\}$, $k = \overline{1, n}$.

По мнению экспертного комитета выходная характеристика качества работы общеобразовательных учреждений определяется, в основном, двумя составляющими учебного процесса: уровнем квалификации преподавательского состава и качеством подготовки выпускников школ в аспекте дальнейшего повышения уровня образования, а также эффективной работы в различных сферах человеческой деятельности.

Соответственно были выбраны две агрегированные выходные переменные: Φ_1 — обобщенный показатель квалификации преподавательского состава; Φ_2 — обобщенный показатель качества полученных знаний.

Показатель качества Φ_1 определим как агрегат из первичных показателей $\{\Phi_{1r}\}$:

$$\Phi_1 = \sum_{r=1}^3 d_r \Phi_{1r}; \quad \sum_{r=1}^3 d_r = 1; \quad d_r \geq 0, \quad (2)$$

где Φ_{11} , Φ_{12} , Φ_{13} — доли работников с высшим профессиональным образованием, с высшей категорией (или ученой степенью), с первой квалификационной категорией в общем числе педагогических работников данного образовательного учреждения.

Будем считать, что наличие у работников образовательных учреждений высшего образования (показатель Φ_{11}) является необходимым условием обеспечения учебного процесса и со временем Φ_{11} будет стремиться к 1, поэтому весовой множитель d_1 выбран наименьшим и равным 0,2.

Показатели Φ_{12} и Φ_{13} , связанные с числом педагогических работников высокой квалификационной категории, — это "спинной хребет" учебного процесса в аспекте усовершенствования существующих методик обучения, а также апробации и

внедрения принципиально новых, инновационных методик и образовательных технологий. Соответственно, весовой коэффициент d_3 выбран равным 0,35, а $d_2 = 0,45$.

Показатели Φ_{12} и Φ_{13} формируются общими усилиями коллектива данного образовательного учреждения и в то же время отражают косвенно (через фонд оплаты труда) политику субъекта Российской Федерации по реализации национальных проектов.

Таким образом, получаем формулу для расчета агрегата Φ_1 :

$$\Phi_1 = 0,2\Phi_{11} + 0,45\Phi_{12} + 0,35\Phi_{13}. \quad (3)$$

Теперь сделаем расчет агрегированного показателя Φ_2 :

$$\Phi_2 = \sum_{r=1}^5 g_r \Phi_{2r}; \quad \sum_{r=1}^5 g_r = 1; \quad g_r \geq 0, \quad (4)$$

где Φ_{21} — показатель качества знаний (отношение числа отличных и хороших оценок к общему числу оценок); Φ_{22} — отношение числа учащихся — призеров городских олимпиад, конкурсов, соревнований к общему числу учащихся данного образовательного учреждения; Φ_{23} — отношение числа учащихся — призеров республиканских и всероссийских олимпиад, конкурсов, соревнований к общему числу учащихся данного образовательного учреждения; Φ_{24} — отношение числа учащихся, поступивших в вузы России к общему числу учащихся данного образовательного учреждения; Φ_{25} — отношение числа выпускников, награжденных медалями "За особые успехи в учении", серебряными и золотыми, к общему числу учащихся данного образовательного учреждения.

Группа экспертов присвоила следующие значения весовых множителей: $\{g_r\}$: $g_1 = g_4 = 0,05$; $g_2 = g_5 = 0,25$; $g_3 = 0,4$.

Оценки экспертов продиктованы следующими соображениями. Прежде всего выделен показатель Φ_{24} — относительного числа учащихся, поступивших в вузы. На первый взгляд, этот показатель носит объективный характер, поскольку он связан с оценками знаний абитуриентов в рамках ЕГЭ. Однако данный показатель не дифференцирует число поступивших на бюджетную и договорную формы обучения, т. е. по существу, не различает социальный статус лиц, оплачивающих обучение. Другими словами, оценка качества подготовки учащихся — абитуриентов на основе показателя Φ_{24} нивелируется различием проходных баллов для бюджетной и договорной форм обучения в вузе. Действительно, из приведенной здесь таблицы видно, что показатель Φ_{24} для девяти объектов, наблюдаемых в течение 5 лет, колеблется от ми-

Фрагмент данных для построения НСМ

Номер объекта S	Сквозной номер наблюдения i	Относительное текущее время наблюдения X_1	X_2	X_3	X_4 , тыс. руб./чел.	X_5 , тыс. руб./чел.	X_6 , тыс. руб./чел.	Φ_1	Φ_2	Φ
1	2	3	4	5	6	7	8	9	10	11
1	1	0,0417	1	0	12,45	3,25	2189	0,364	0,0634	0,0231
	2	0,0833	1	0	13,1	3,25	2295	0,364	0,0634	0,0231
	3	0,125	1	0	3,7	3,25	653,4	0,364	0,0634	0,0231
	4	0,166	1	0	17,2	3,25	3030	0,364	0,0634	0,0231
	5	0,208	1	0	12,7	4,38	2297	0,285	0,0678	0,0193
	6	0,25	1	0	15,2	4,38	2749	0,285	0,0678	0,0193
	7	0,292	1	0	3,97	4,38	718,3	0,285	0,0678	0,0193
	8	0,333	1	0	17,2	4,38	3104	0,285	0,0678	0,0193
	9	0,375	1	0	12,6	5,21	2394	0,614	0,0678	0,0193
	10	0,417	1	0	14,5	5,21	2743	0,614	0,0678	0,0193
	11	0,458	1	0	4,03	5,21	765,7	0,614	0,0678	0,0193
	12	0,5	1	0	20,0	5,21	3789	0,614	0,0678	0,0193
	13	0,542	1	0	13,2	5,98	2619	0,313	0,0692	0,0216
	14	0,583	1	0	15,9	5,98	3143	0,313	0,0692	0,0216
	15	0,625	1	0	3,7	5,98	733,4	0,313	0,0692	0,0216
	16	0,666	1	0	20,1	5,98	3981	0,313	0,0692	0,0216
	17	0,708	1	0	14,1	6,78	2893	0,325	0,0682	0,0222
	18	0,75	1	0	16,4	6,78	3360	0,325	0,0682	0,0222
	19	0,792	1	0	4,1	6,78	831,4	0,325	0,0682	0,0222
	20	0,833	1	0	21,0	6,78	4305	0,325	0,0682	0,0222
2	21	0,0417	1	0	12,55	3,26	2110	0,264	0,064	0,0169
	22	0,0833	1	0	15,1	3,26	2532	0,264	0,064	0,0169
	23	0,125	1	0	4,52	3,26	759,7	0,264	0,064	0,0169
	24	0,166	1	0	18,1	3,26	3039	0,264	0,064	0,0169
...	...	...	...	...	...	...	...	...	...	...
9	180	0,833	0	1	23,79	36,78	5381	0,317	0,052	0,0165

нимального значения 0,57 до максимального 0,7, т. е. всего на 18,57 % (см. прилож. 5 из работы [3]).

Показатель Φ_{21} качества обучения отягчен требованиями вышестоящих органов по обеспечению качества успеваемости и потому неизбежно создает у преподавателей стимул "натягивания" оценок по итогам года, т. е. вносит элемент нивелирования.

Таким образом, показатели Φ_{21} и Φ_{24} малоинформативны в аспекте оценки истинного уровня качества получаемых знаний, поэтому им присвоены малые весовые коэффициенты $d_1 = d_4 = 0,05$.

Более объективно характеризуют создание в образовательном учреждении благоприятных условий для усвоения образовательных программ и развития талантов и личности учащихся показатели Φ_{22} , Φ_{23} и Φ_{25} , связанные с числом медалистов и призеров различных олимпиад и конкурсов городского уровня. Соответственно весовые коэффициенты выбраны равными $d_3 = 0,4$, $d_2 = d_5 = 0,25$.

В итоге получим формулу линейной свертки для агрегата Φ_2 :

$$\Phi_2 = 0,05\Phi_{21} + 0,25\Phi_{22} + 0,4\Phi_{23} + 0,05\Phi_{24} + 0,25\Phi_{25}. \quad (5)$$

Для получения обобщенного показателя эффективности работы образовательного учреждения, являющегося выходной величиной нейросетевой модели, формируется мультипликативная свертка из агрегатов Φ_1 и Φ_2 :

$$\Phi = \Phi_1\Phi_2. \quad (6)$$

Теперь опишем спецификацию независимых переменных X . Для планирования муниципального бюджета удобно, чтобы время t входило в модель явно, поэтому первая независимая переменная X_1 — это относительное время:

$$X_1 = t/T, \quad (7)$$

где T — период наблюдения; t — текущее время (месяцы).

К сожалению, введение лаговых переменных, учитывающих предысторию процесса ("медленное время"), для данных из таблицы невозможно, поскольку выходные показатели качества, в силу своей специфики, измеримы только раз в год. В базе данных они условно пролонгированы по кварталам.

Оценки экспертного комитета, а также предобработка данных с помощью итерационной процедуры оптимальной кластеризации из работ [2, 3, 5]

показали, что можно выделить три кластера МУ, для которых значения обобщенного показателя Φ в свертке (6) существенно отличаются. Средние значения $\bar{\Phi}$ для кластеров 1, 2 и 3 соответственно равны $\bar{\Phi}^{(1)} = 0,023$; $\bar{\Phi}^{(2)} = 0,017$; $\bar{\Phi}^{(3)} = 0,015$, т. е., если за 100 % принять $\bar{\Phi}^{(1)}$, то $(\bar{\Phi}^{(2)}/\bar{\Phi}^{(1)})100\% = 73,9\%$, а $(\bar{\Phi}^{(3)}/\bar{\Phi}^{(1)})100\% = 65,2\%$.

Таким образом, выявлено заметное "расслоение" образовательных учреждений, которое хорошо согласуется с качественными представлениями в этой области: инновационные образовательные учреждения (лицей или гимназия) имеют наилучшие качественные показатели по квалификационному уровню преподавателей и по качеству успеваемости, наличию медалистов и призеров олимпиад и конкурсов высшего уровня, а школы-интернаты — наихудшие показатели. Обычные школы занимают среднее положение.

Поскольку имеет место дефицит наблюдений (в таблице всего 216 примеров) и данные сильно зашумлены, авторы приняли решение не разбивать их на кластеры, а упомянутый эффект "расслоения" учесть введением двух бинарных переменных:

$$X_2 = \begin{cases} 1, & \text{если объект — лицей и гимназия;} \\ 0, & \text{в противном случае.} \end{cases}$$

$$X_3 = \begin{cases} 1, & \text{если объект — школа-интернат;} \\ 0, & \text{в противном случае.} \end{cases}$$

Заметим, что если увеличить число бинарных переменных, то можно расширить номенклатуру моделируемых объектов, т. е. рассмотреть, например, музыкальные образовательные учреждения и др.

Приведем главные эвристические соображения по выбору остальных входных факторов, которые фигурировали в указанной выше процедуре экспертных оценок с "дообучением экспертов".

Безусловно, фонд оплаты труда (ФОТ) и начисления на оплату труда учителей и обслуживающего персонала напрямую связаны с квалификацией учителей (показателем Φ_1) и косвенно с показателем Φ_2 , характеризующим качество знаний. Поэтому был введен входной фактор X_4 — относительное значение фонда оплаты труда и начислений на фонд оплаты труда, приходящийся на одного работника в течение квартала.

Уровень материального обеспечения учебного процесса по всей видимости также оказывает влияние на усвоение учебного материала, поскольку создает хотя бы минимальный комфорт в классах, обеспечивает образовательные технологии вычислительной техникой, приборами, учебными пособиями и т. д. Немаловажную роль играют расходы на проведение конкурсов, олимпиад, смотров, экскурсий.

Уровень материального обеспечения учебного процесса характеризуется в основном двумя фак-

торами: X_5 — расход средств на содержание одного ученика в течение года, тыс. руб./чел.; X_6 — общая сумма расходов данного образовательного учреждения за квартал, тыс. рублей.

Заметим, что фактор X_6 характеризует масштаб деятельности учреждения. Доходы учреждений от предпринимательской деятельности несущественны (составляют примерно 4 % от бюджетных расходов) и практически одинаковы для всех девяти объектов из рассматриваемого кластера. Данный фактор не будем вводить в модель, поскольку он неинформативен.

Фрагмент исходных данных в соответствии с описанной методикой спецификации переменных приведен выше в таблице. Период наблюдения составлял 5 лет, общее число наблюдений 216.

Перед обучением НСМ проводилась процедура предобработки — итерационная оптимальная очистка данных от аномальных точек. Соответствующая задача оптимизации, целью которой является повышение однородности данных, имеет вид:

$$\left\{ \begin{array}{l} i^{*(k+1)}: \left[\max_{i \in N^{(k)}} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| 100\% \right] > \eta_1, \\ J(N^{(k)}, E^{(k)}) \rightarrow \min_k J; k = 0, 1, 2, \dots, N^{(k)} \geq \xi n; \\ J = C_1 E + C_2 (NMSE); \\ E: i \in \Omega_{test}; NMSE: i \in \Omega_{learn} \end{array} \right. \quad (8)$$

где $i^{*(k+1)}$ — номер точки, которая на следующей $(k+1)$ -й итерации отбраковывается, т. е. выводится из данных как аномальная; $N^{(k)}$ — число точек в данных на k -й итерации в обучающем множестве; η_1 — уровень относительной ошибки, по которому проводится отбор аномальных точек; J — векторный критерий минимизации, зависящий от числа точек $N^{(k)}$ в данных на k -й итерации и ошибки обобщения $E^{(k)}$; ξ — экспериментально подбираемый коэффициент репрезентативности выборки ($\xi \approx 5 \dots 10$); n — число входных факторов; C_1, C_2 — весовые множители, задаваемые экспертно (принято $C_1 = C_2 = 0,5$); $NMSE$ — нормированная средняя квадратическая ошибка расчета по НСМ; $\Omega_{learn}, \Omega_{test}$ — обучающее и тестовые множества.

Состоятельность задачи векторной (многокритериальной) оптимизации (8) обусловлена теми же соображениями, что и задачи оптимальной кластеризации [2, 3]. При пошаговой отбраковке аномальных точек в каждом из образованных кластеров растет однородность данных в кластере по евклидовой норме и, как следствие — снижается ошибка обобщения E и векторный критерий J в

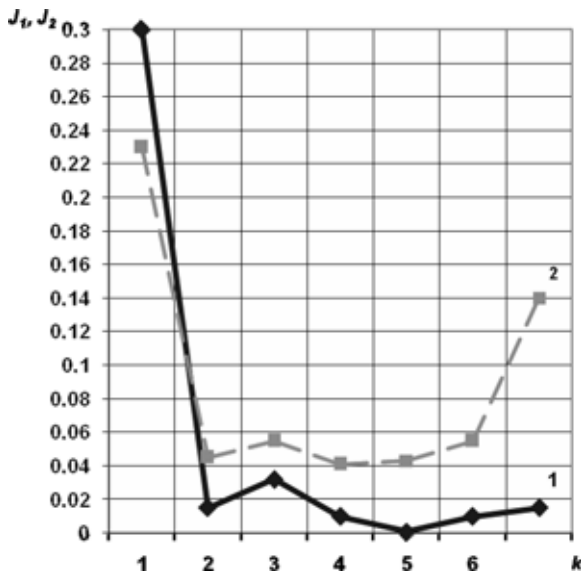


Рис. 2. Результаты оптимальной очистки данных:
1 — минимизация по критерию J_1 ; 2 — по критерию J_2

системе (8). Однако при этом снижается репрезентативность в выборке, измеряемая числом N^k оставшихся после удаления аномальных точек примеров в кластере. Это вызывает ухудшение (рост) векторного критерия (8). Поэтому существует оптимальный номер итерации k^* и, соответственно, число наблюдений $N^{*(k)}$ в "очищенном" кластере, доставляющий минимум векторному критерию J . Замечательно, что минимум J достигается уже на четвертой — пятой итерации (рис. 2).

На рис. 2 показана зависимость векторного аддитивного критерия минимизации J по (8) от номера итерации k процедуры очистки кластера и, соответственно, числа оставшихся в работе точек $N^{(k)}$. Заметим, что первое слагаемое в критерии J_1 — ошибка обобщения E в (8) характеризует среднюю квадратическую ошибку аппроксимации сети на точках, которые HCM ранее не "знала" (они не участвовали в обучении), а второе слагаемое $NMSE$ — это средняя квадратическая ошибка аппроксимации уже обученной HCM на всем множестве точек $N^{(k)}$ на k -й итерации. На том же рисунке показаны результаты этой процедуры по другому, мультипликативному критерию минимизации:

$$J_2 = E \cdot NMSE, \quad (9)$$

который не содержит экспертно задаваемых весовых коэффициентов.

Из рис. 2 видно, что оба критерия дают примерно одинаковый результат: оптимальный кластер после очистки соответствует итерации $k = 5$. Причем уменьшение ошибки обобщения E на первых трех итерациях составило в 2,61 раз, а $NMSE$ в 6,18 раз. Критерии минимизации уменьшились J_1 в 5,13 раз и J_2 в 169 раз. Небольшой "всплеск" кривых $J_{1(k)}$ и $J_{2(k)}$ в точке $k = 2$ объясня-

ется тем, что в этой точке HCM было изменено число нейронов во втором скрытом слое (вместо $n_2 = 6$ было взято $n_2 = 3$). Качество очистки оказалось чувствительным к вариации этого параметра n_2 .

Минимум на рис. 2 выражен очень четко. Вначале на двух итерациях происходит резкое уменьшение E , $NMSE$, а также аддитивной и мультипликативной свертки этих показателей, затем после итерации $k = 4$ и удалении из БД более 8 % точек указанные показатели начинают возрастать. Это объясняется нарастающим снижением репрезентативности данных, т. е. нарушением ограничения в (8).

Можно сделать общий вывод: предложенная процедура оптимальной итерационной очистки кластера от аномальных точек по критериям E и $NMSE$, а также по векторным критериям J_q (E , $NMSE$), $q = 1, 2$, оказалась весьма эффективным инструментом предобработки данных. Ценой небольшого уменьшения объема БД (со 180 точек до 165 точек, т. е. на 8,3 %) уже на третьей итерации удалось резко улучшить качество HCM. В дальнейших расчетах использовался очищенный кластер Ω^* .

3. Оптимизация архитектуры сети. Оценка адекватности модели и результаты моделирования

Серия вычислительных экспериментов [3] показала, что в рассматриваемой задаче качество HCM существенно зависит от архитектуры сети и вида активационных функций, поэтому решалась соответствующая нелинейная задача параметрической оптимизации. Фиксировался тип HCM — многослойный персептрон (MLP) с обратным распространением ошибки при обучении. Параметры сети: число скрытых слоев u_1 , которое варьировалось от 1 до 2; число нейронов в первом скрытом слое u_2 изменялось от 1 до 20; число нейронов во втором скрытом слое u_3 принимало значение от 1 до 20; тип активационной функции u_4 — бинарная переменная; $u_4 = 1$, если $f(s)$ — сигмоидная функция, $u_4 = 0$, если $f(s)$ — гиперболический тангенс. В качестве критерия качества аппроксимации сети использовался векторный критерий минимизации $Q = (Q_1, Q_2, Q_3, Q_4, Q_5)$, где Q_1 — финальная средняя квадратическая ошибка (НМСЕ) на обучающем множестве Ω^{train} ; Q_2 — ошибка обобщения (НМСЕ на тестовом множестве Ω^{test}); Q_3 — максимальная относительная ошибка на Ω^{test} ; Q_4 — коэффициент риска выхода относительной ошибки прогноза за назначенный предел [3], Q_5 — показатель качества аппроксимации, $Q_5 = [1 - (r_Y, \hat{y})^2]$.

Находим минимум по мультипликативной свертке $Q = \prod_{v=1}^5 Q_v(u)$ в два этапа: на первом этапе

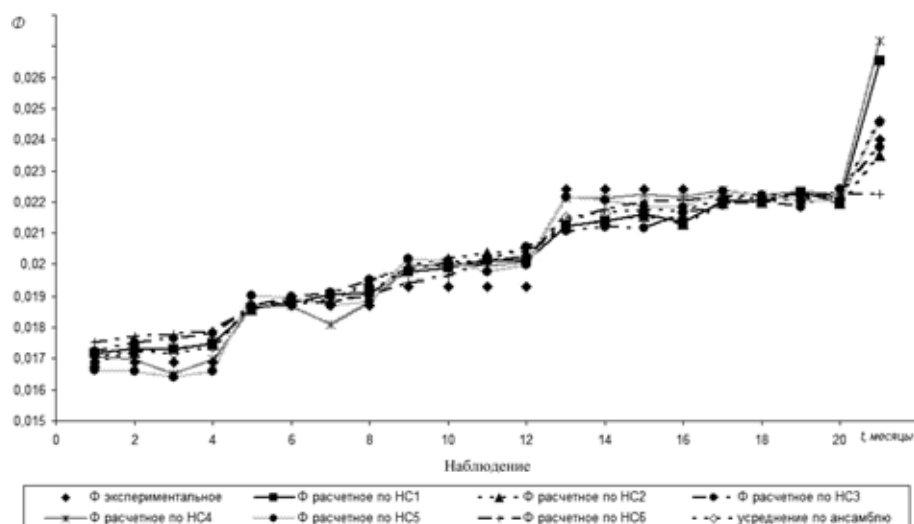


Рис. 3. Прогноз Φ по объекту 2

использовался приближенный псевдослучайный поиск с помощью зондирования многофакторного пространства точками ЛП_т-последовательностей И. М. Соболя [6]. Найденное оптимальное значение u^* на втором этапе использовалось как начальное приближение для задачи целочисленного программирования, которая решалась методом ветвей и границ [7].

Результаты оптимизации: число скрытых слов — 1; число нейронов в первом слое — 5; число нейронов во втором слое — 2; вид активационной функции — сигмоид. Рекомендуемый тип сети — MLP с алгоритмом обучения BP [4], алгоритм обработки примеров — пакетный, коэффициент обучения был взят для всех точек равным 0,1. Использовался программный продукт NeuroSolutions (демоверсия). Число эпох обучения — 10000. Эффекта переобучения HSM не обнаружено.

Адекватность HSM оценивали по методу обобщенного перекрестного подтверждения [3]. Строили шесть параллельных (независимых) сетей с различными архитектурами и активационными функциями в рамках описанных выше вычислительных экспериментов по оптимизации. Результаты показаны на рис. 3, из которого видно, что расчетные кривые для шести сетей (HSM1...HSM6) хорошо согласуются между собой. Разброс прогнозных значений $\Phi_{2\text{прогн}}$ для объекта 2 (школы-интернаты) составляет 0,005 или по отношению к $\bar{\Phi}$ 20 %, что можно считать неплохим результатом для социально-экономических объектов с сильным зашумлением данных [3]. Для других восьми объектов прогнозные кривые для шести параллельных моделей имеют аналогичный характер.

Общие результаты оценки коэффициента эффективности для девяти общеобразовательных средних МУ: $\Phi_1 = 0,0245$; $\Phi_2 = 0,0246$; $\Phi_3 = 0,0231$;

$\Phi_4 = 0,0225$; $\Phi_5 = 0,0224$; $\Phi_6 = 0,0228$; $\Phi_7 = 0,0233$; $\Phi_8 = 0,0136$; $\Phi_9 = 0,0138$.

Здесь объекты 1 и 2 — школы-лицеи; 3...7 — обычные средние школы; 8 и 9 — школы-интернаты [3].

Таким образом, вычислительные эксперименты в нейросетевом базисе подтвердили экономическую концепцию, сформулированную в начале статьи.

Выводы

1. На основе нейросетевой модели предложена экономико-математическая концепция "замыкания" снизу (на уровне МУ) нормативной методики [1] распределения субвенций в процессах муниципального бюджетирования.

2. При построении HSM использованы оригинальные процедуры предобработки данных: экспертных оценок с "дообучением" экспертов и очистки образованных кластеров данных. Отличительная черта этих процедур — связь с качеством обучения сети.

3. Все предложения по построению HSM апробированы в вычислительных экспериментах на реальных данных.

4. По мнению авторов, развитый в статье подход применим для широкого круга задач финансового контроля, где требуется ранжирование экономических объектов [1—3, 5, 8].

Список литературы

1. Методические рекомендации субъектам Российской Федерации и муниципальным образованиям по регулированию межбюджетных отношений // Приказ Министерства финансов РФ № 243 от 27 августа 2004 г.
2. Горбатков С. А., Полупанов Д. В. Методы нейроматематики в налоговом контроле. Уфа: РИЦ Баш. ГУ, 2008.
3. Гатауллин Р. Ф., Горбатков С. А., Бирюков А. Н., Глущенко О. И. Моделирование бюджетных процессов на муниципальном уровне на основе нейросетей. Уфа: Восточный университет, 2008.
4. Галушкин А. И. Теория нейронных сетей. Кн. 1: Учеб. пособие для вузов / Общая ред. А. И. Галушкина. М.: ИПРЖР, 2000. 416 с.
5. Бублик Н. Д., Голичев И. И., Горбатков С. А., Смирнов А. В. Теоретические основы разработки технологии налогового контроля и управления. Уфа: РИО БашГУ, 2004.
6. Соболев И. М., Статников Р. Б. Выбор оптимальных параметров в задаче со многими критериями. М.: Наука, 1981.
7. Жигляевский А. А., Жилинскас А. Г. Методы поиска глобального экстремума. М.: Наука, 1991.
8. Бэстенс Д.-Э., Ван дер Берг В.-М., Вуд Д. Нейронные сети и финансовые рынки: принятие решений в торговых операциях. — М.: ТВП, 1998.

УДК 004.3.001.57

С. С. Соболев¹, аспирант,
Л. С. Сотов², канд. физ.-мат. наук, доц.,
e-mail: slskit@mail.ru,

В. Н. Харин¹, д-р техн. наук, проф.

¹ Воронежская государственная
лесотехническая академия

² Саратовский государственный университет

Алгоритм работы и модель функционального генератора перестановок

Предложены алгоритм работы и модель высокопроизводительного генератора перестановок чисел $(0, 1, \dots, n)$. Генератор имеет матричную структуру и аппаратную сложность $C_N = \frac{n}{2} (\log_2^2(n) - \log_2(n))$ условных логических элементов, что дает возможность формирования перестановок при больших значениях n . Предложенный в работе генератор перестановок предназначен для системы аппаратного форматирования бинарных строк представления информации в ЭВМ.

Ключевые слова: генератор перестановок, генератор сочетаний, баньян-переключатель, динамические перестановки, псевдослучайные перестановки

Введение

Разработка и исследования быстродействующих генераторов перестановок представляют интерес в различных приложениях, в том числе устройствах аппаратной поддержки технологий систем автоматизированного проектирования радиоэлектронной аппаратуры [1]. В технике коммуникаций с использованием псевдослучайного переключения частот использование генераторов перестановок может повысить равномерность использования частот, так как при этом за некоторый промежуток времени используются все частоты выделенного диапазона [2]. В системах защиты информации используются различные шифры, основанные на перестановках [3, 4].

Исследованиям генераторов и устройств осуществления перестановок посвящено достаточно много работ, обзор многих из которых можно найти в [1]. В ряде работ рассмотрено расширение архитектуры процессоров для быстрого осуществ-

ления операций перестановки, например [5]. Среди генераторов перестановок особое место занимают функциональные генераторы (формирователи), которые реализуют заданную перестановку по ее номеру [1].

Важнейшими параметрами, характеризующими генераторы перестановок, являются простота аппаратной реализации, производительность, простота управления и настройки. Оптимизация по данным параметрам противоречива, поэтому для конкретных задач определяется компромиссное решение.

Наиболее быстрыми оказываются устройства, осуществляющие параллельную перестановку всех элементов [6]. Их аппаратная сложность растет пропорционально n^2 , где n — число переставляемых чисел. Оптимальная аппаратная сложность достигается в устройствах, работа которых основана на сетях Бенеша [7]. Аппаратная сложность в этом случае растет как $n \log_2 n$, т. е. практически линейно. Однако перестановка в этом случае осуществляется последовательно-параллельно за $2 \log_2 n$ преобразований на уровнях сети, и это снижает быстродействие в соответствующее число раз. Кроме этого, управление перестановками, связанное с настройкой сети Бенеша, достаточно трудоемкое, несмотря на усилия, предпринимаемые исследователями в этом направлении [8].

В данной работе предложена модель быстродействующего функционального генератора перестановок чисел $(0, 1, \dots, n)$, отличительными особенностями которого являются простота управления, возможность формирования перестановки по ее номеру и достаточно высокая производительность в режиме псевдослучайной генерации. Аппаратная сложность генератора перестановок растет как

$$C_N = \frac{n}{2} (\log_2^2(n) - \log_2(n)).$$

Данный генератор предназначен для аппаратного обеспечения динамического преобразования форматов представления данных в ЭВМ [9].

Алгоритм формирования перестановок

В основе предложенного алгоритма формирования перестановок чисел строки $(0, 1, \dots, n)$ лежит последовательное использование процедуры поразрядного формирования сочетаний битов,

Двоичное представление чисел

Число	Разряды		
	3	2	1
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

входящих в соответствующий двоичный разряд переставляемых чисел. В таблице приведены числа (0, 1, ..., 7), представленные в двоичной системе счисления.

В столбце, в котором представлены первые разряды чисел строки (0, 1, ..., 7), расположены четыре пары битов, образующих бинарные строки (0,1). Каждая пара имеет одинаковые старшие разряды в соответствующей строке таблицы. На первом шаге предложенного алгоритма формируются четыре сочетания битов $s_{11}(0,1)$, $s_{12}(0,1)$, $s_{13}(0,1)$, $s_{14}(0,1)$. Общее число возможных комбинаций, образуемых данными сочетаниями битов, составляет $\binom{2}{1}^4 = \frac{2!}{(1!)^2} = 2^4 = 16$.

Рассмотрим столбец, содержащий биты второго разряда. Эти биты можно объединить в две строки $S_1 = (0, 0, 1, 1)$, $S_2 = (0, 0, 1, 1)$. Каждый элемент в составе S_1 и S_2 расположен в таблице так, что имеет одинаковые старшие разряды в соответствующей строке таблицы. На втором шаге алгоритма формируются два сочетания битов $s_{21}(0, 0, 1, 1)$, $s_{22}(0, 0, 1, 1)$. Число возможных комбинаций, образуемых данными сочетаниями битов, составляет $\binom{4}{2}^2 = 6^2 = 36$.

Число возможных комбинаций, образуемых сочетаниями битов бинарной строки (0, 0, 0, 1, 1, 1, 1, 1), составляет $\binom{8}{4} = 70$. Таким образом, число перестановок, формируемых в результате работы данного алгоритма, составляет $C_8 = \binom{2}{1}^4 \binom{4}{2}^2 \binom{8}{4} = 8!$ В общем случае

$$C_n = \prod_{i=1}^{\log_2 n} \binom{n/2^{i-1}}{n/2^i} = n!$$

На рис. 1 приведена диаграмма орграфа, иллюстрирующая приведенный выше алго-

ритм формирования перестановки строки (0, 1, ..., 7). Перестановка генерируется в три этапа. На каждом этапе осуществляется формирование сочетаний битов соответствующего разряда и добавление битов старшего разряда.

Модель генератора перестановок

Генератор перестановок построен с использованием формирователей сочетаний элементов бинарных строк, в которых число битов с высоким логическим уровнем равно числу битов с низким логическим уровнем.

Генератор перестановок построен по матричной схеме. На рис. 2 представлена его блок-схема для случая $n = 8$ элементов. Матрица состоит из трех уровней преобразования $J = 1, 2, 3$. На первом уровне располагаются четыре блока промежуточного преобразования. Каждый из блоков первого уровня реализует перестановку элементов бинарной строки (0, 1). Таким образом, на первом уровне преобразования формируются четыре перестановки бинарных строк (0, 1).

На втором уровне преобразования в каждом блоке преобразования к каждому биту сформированных перестановок добавляется старший разряд из входной бинарной строки данных (0, 0, 1, 1). Причем два первых добавленных бита имеют высокий логический уровень. Таким образом, в блоках второго уровня преобразованию подвергаются числа {0, 1, 2, 3}. Преобразование заключается в том, что в блоках второго уровня формируются сочетания входной бинарной строки данных (0, 0, 1, 1), причем "прикрепленный" бит младшего разряда также переставляется вместе с битом старшего разряда. Таким образом, на втором уровне преобразования формируются две перестановки строк (0, 1, 2, 3). Перестановка строки (0, 1, 2, 3, 4, 5, 6, 7) формируется на выходе третьего уровня преобразования.



Рис. 1. Диаграмма орграфа формирования перестановки, $n = 8$

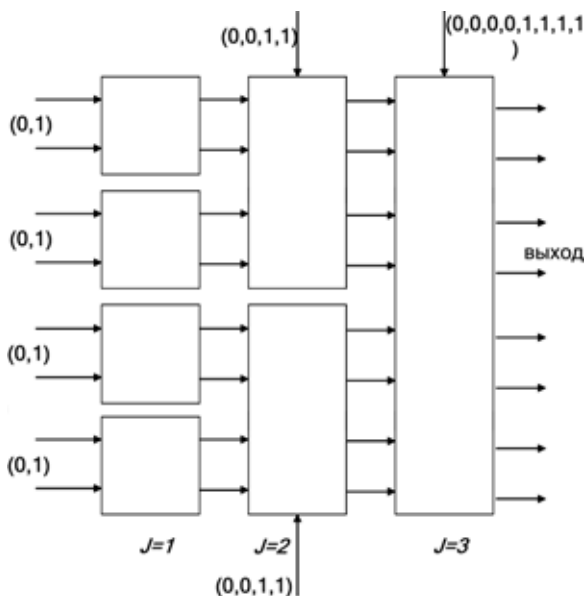


Рис. 2. Блок-схема функционального генератора перестановок, $n = 8$

В общем случае на каждый из блоков формирования перестановки длиной n элементов поступают две перестановленные строки длиной $n/2$ элементов. Число различных комбинаций на входах блока составляет $N_1 = ((n/2)!)^2$. Число различных сочетаний элементов входной бинарной строки составляет $N_2 = \binom{n}{n/2} = \frac{n!}{((n/2)!)^2}$. Таким образом, общее число перестановок, которое реализуется на выходе блока, составляет $N_1 N_2 = n!$

Для формирования сочетаний элементов бинарных строк предлагается метод, иллюстрируемый диаграммой графа $G(S, D, T, A)$, представленной на рис. 3, где $S = (s_0, s_1, s_2, \dots, s_i, \dots, s_n - 1)$ — линейно упорядоченное множество исходных вершин графа для задания битов сочетания, имеющих высокий и низкий логические уровни; $D = (d_0, d_1, d_2, \dots, d_i, \dots, d_n - 1)$ — линейно упорядоченное множество вершин назначения графа, которое можно рассматривать как выходы генератора сочетаний; T — множество промежуточных вершин графа; A — множество ребер графа.

Вершины $t_{ij} \in T$ можно записать в виде матрицы, имеющей $n/2$ строк, $k = \log_2 n$ столбцов — уровней преобразования. Каждая вершина t_{im} уровня $m = (1, k - 1)$ является смежной двум вершинам t_{hm+1}, t_{pm+1} уровня $m + 1$, причем на значения индексов h и p наложены ограничения

$$2^{k-m-1} \text{int} \left(\frac{i-1}{2^{k-m-1}} \right) + 1 \leq h \leq 2^{k-m-1} \text{int} \left(\frac{i-1}{2^{k-m-1}} + 1 \right); \quad (1)$$

$$2^{k-m-1} \text{int} \left(\frac{(i + 2^{k-m-1} - 1) \bmod n}{2^{k-m-1}} \right) + 1 \leq p \leq 2^{k-m-1} \left(\text{int} \left(\frac{(i + 2^{k-m-1} - 1) \bmod n}{2^{k-m-1}} \right) \right),$$

где int — функция выделения целой части; $(j + 2^{k-m-1} - 1) \bmod n$ — функция вычисления остатка от частного $\frac{j + 2^{k-m-1} - 1}{n}$.

Любая вершина уровня $m + 1$ является смежной двум вершинам уровня m . Каждая из вершин множества S является смежной одной из вершин t_{i1} , причем любая из вершин t_{i1} является смежной двум вершинам множества S .

Транспозиционные элементы всех уровней разбиты на подмножества, номер подмножества d определяется выражением

$$d = \text{int} \left(\frac{i + 2^{k-m-1} - 1}{2^{k-m-1}} \right).$$

Учитывая (1), вершины каждого подмножества уровня m можно считать смежными вершинам из двух различных подмножеств уровня $m + 1$, которые далее будем называть верхним и нижним подмножествами.

Для доказательства возможности формирования любого сочетания покажем возможность сортировки с использованием данного графа любой бинарной строки, состоящей из $n/2$ битов с высоким логическим уровнем и $n/2$ битов с низким логическим уровнем.

Пусть $r_{ij} = \bigcup_{l=1}^k \bar{A}_l$ — простой путь от d_j к s_i . Рассмотрим множество путей из D в S

$$R = \{r_{ij} | \forall r_{ij} \neq r_{kl} \Rightarrow r_{ij} \cap r_{kl} = \emptyset\},$$

где R — определяет биективное отображение $T_R: D \rightarrow S$.

Пусть произвольное P — подмножество вершин множества D , которое будем интерпретировать как биты бинарной строки, имеющие высокий логический уровень, и пусть $|P| = n/2$. Рассмотрим множества $C_1 = (s_1, \dots, s_{n/2})$,

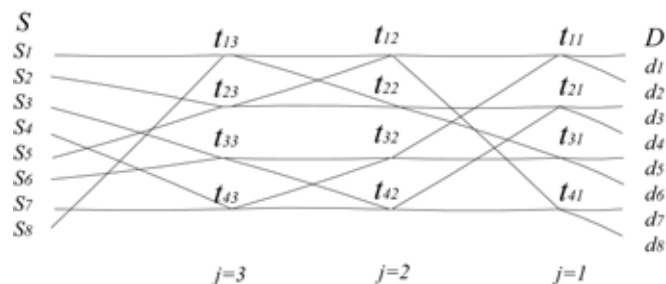


Рис. 3. Диаграмма графа генератора сочетаний, $n = 8$

$C_2 = (s_{n/2+1}, \dots, s_n)$. Для всех $d_i \in P$ существуют $n/2$ различных путей $r_{ij}^1 \in R_1$, $R_1 \subset R$ таких, что $d_j \in C_1$. Для формирования данных путей при переходе между уровнями m и $m+1$ графа $G(S, D, T, A)$ будем выбирать ребра так, чтобы $l/2$ вершин подмножества уровня m соединялись с промежуточными вершинами верхнего подмножества уровня $m+1$, а оставшиеся $l/2$ вершин подмножества уровня m соединялись с промежуточными вершинами нижнего подмножества уровня $m+1$, где l — мощность подмножества уровня m .

Такой выбор всегда можно осуществить, причем выбор ребер участка пути, проходящего через одну из вершин каждого верхнего или нижнего промежуточного подмножества уровня m , можно осуществить произвольно. Таким образом, с использованием описанной выше процедуры можно реализовать сортировку произвольной бинарной строки длиной n . При обратном преобразовании $T_R^{-1} : S \rightarrow D$, задав множества C_1, C_2 , можно сформировать любую бинарную строку, имеющую $n/2$ битов с высоким логическим уровнем и $n/2$ битов с низким логическим уровнем, т. е. любое сочетание.

Число различных отображений $T_R^{-1} : S \rightarrow D$ составляет $\frac{n}{2} \log_2(n)$ и $\frac{n}{2} (\log_2(n) - 1) + l$ при исключении из графа вершин каждого верхнего или нижнего промежуточного подмножества уровня m , пути из которых определяются произвольно.

Архитектурная модель генератора перестановок

Для исследования генератора была разработана архитектурная RTL модель системы с использованием языка SystemC [10]. При проведении исследований использовалась программная среда CoCentric System Studio [11].

В соответствии с рис. 2 генератор перестановок для случая $n = 8$ включает в себя генератор сочетаний элементов бинарной строки длиной 8, два генератора C_4 для формирования сочетаний элементов бинарных строк длиной 4 и четыре генератора сочетаний элементов бинарных строк длиной 2. Схема генератора сочетаний элементов би-

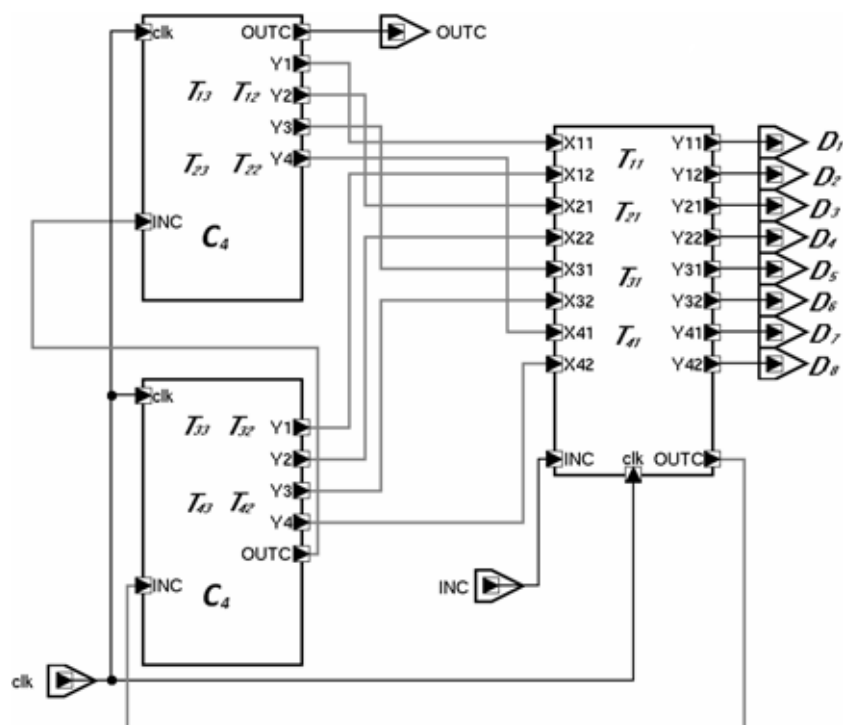


Рис. 4. Схема транспозиционной матрицы генератора сочетаний, $n = 8$

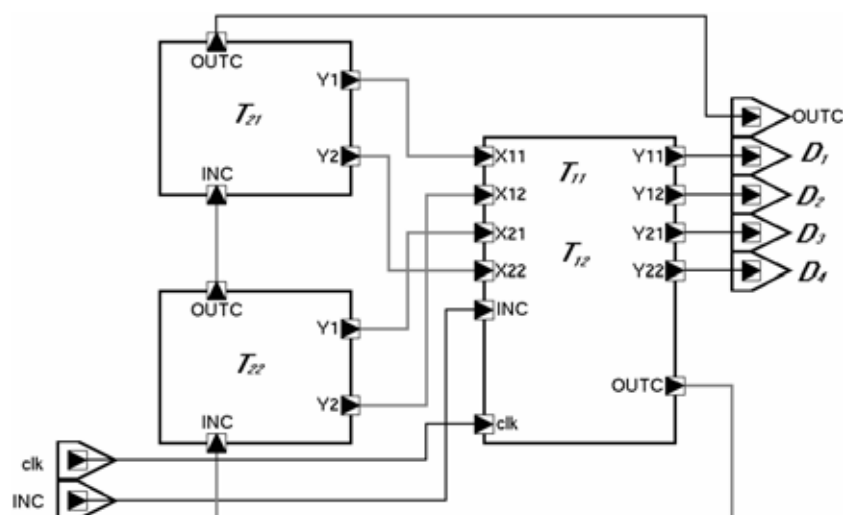


Рис. 5. Схема транспозиционной матрицы генератора сочетаний, $n = 4$

нарной строки длиной 8 представлена на рис. 4. Данная схема включает в себя два генератора C_4 . Схема блока C_4 представлена на рис. 5. Блок состоит из логических транспозиционных элементов T_{ij} , имеющих два входа и два выхода. На элементы T_{ij} последнего уровня преобразования подаются сигналы (0,1) с неизменным логическим уровнем, поэтому на рисунке показаны только выходы Y1, Y2 транспозиционных элементов T_{21}, T_{22} . Элемент T_{11} имеет входы X11, X12 и выходы Y11, Y12. Элемент T_{12} имеет входы X21, X22 и выходы Y21, Y22.

Элементы T_{ij} управляются внутренним сигналом COD (на рисунке он не показан), поступающим от встроенного счетного триггера. Если этот сигнал имеет высокий логический уровень, элемент T_{ij} осуществляет транспозицию поступающих на входы данных, в противном случае данные с входов на выходы передаются без транспозиции. Таким образом, каждый логический элемент T_{ij} представляет собой Баньян-переключатель 2×2 (Banyan switch).

Соединения между транспозиционными элементами T_{ij} выполнены в соответствии с правилами, определенными при описании отношения смежности вершин t_{ij} графа, представленного на рис. 3. Счетные триггеры образуют двоичный счетчик длиной $\frac{n}{2}(\log_2(n) - 1) + 1$. На один из вхо-

дов блока преобразования первого уровня подается логическая 1, на второй — логический 0. Вход INC и выход OUTC модели являются тактовыми входом и выходом счетных триггеров. Счет происходит по заднему фронту сигнала на входе INC. Для ускорения работы генератора счетчик на каждом уровне преобразования может строиться по синхронной схеме. На вход clk подаются тактовые импульсы генератора. По переднему фронту clk бит переноса OUTC устанавливается в состояние с высоким логическим уровнем.

Рассмотрим ограничения, накладываемые на управляющие сигналы COD. Если на входы транспозиционного элемента поступают сигналы с одинаковым логическим уровнем, выходные сигналы не зависят от управляющего бита. Будем называть такой случай исключением управляющего сигнала или просто исключением. Заметим, что исключения на уровне преобразования возникают парами, т. е. их число всегда четное. С учетом исключений длина управляющего кода составляет

$$m_1 = m - \sum_{i=2}^k b_i,$$

где b_i — число исключений на уровне преобразования. На последнем уровне $j = 3$ (см. рис. 3) исключений не возникает. Рассмотрим уровень $j = 1$. На один из входов транспозиционных элементов подается сигнал от транспозиционного элемента верхнего подмножества уровня $j = 2$, на другой — от транспозиционного элемента нижнего подмножества уровня $j = 2$. Число сочетаний, которое генерирует каждое верхнее и нижнее подмножество транспозиционных элементов, составляет $\binom{4}{2} = 6$.

Всего комбинаций 36. Из них — шесть вариантов выбора столбцов с одинаковым номером при-

водит к числу исключений, равному 4, шесть вариантов без исключений и 24 варианта с двумя исключениями. Заметим, что варианты с различной структурой исключений дают в итоге различные сочетания. Из 24 вариантов с двумя исключениями следует оставить 12, так как от перемены порядка верхнего и нижнего подмножеств структура исключений на уровне $j = 1$ не изменяется. Пусть множество $X = \{0, 0, 1, 1\}$. Выбираем из множества произвольно два элемента. Число способов

выбрать $\{1, 1\}$ или $\{0, 0\}$ равно $\binom{4}{2} + \binom{4}{2} = 12$. Из шести вариантов без исключений следует оставить один, так как остальные варианты дадут аналогичные сочетания.

Учитывая, что в варианте без исключений число управляющих битов 4, в варианте с двумя исключениями число управляющих битов 2, а в варианте с четырьмя исключениями число управляющих битов 0, число различных сочетаний, формируемых матрицей, составляет

$$6 \cdot 2^0 + 12 \cdot 2^2 + 1 \cdot 2^4 = 70.$$

Сформулируем *правила запрещенных состояний триггеров счетчика*.

1. Если логические уровни сигналов на входах транспозиционного элемента одинаковые, уровень COD триггера может иметь только низкий логический уровень.

2. Если логические уровни сигналов на входах всех транспозиционных элементов уровня преобразования различные, то допустимыми являются только две комбинации, при которых на входы неуправляемого транспозиционного элемента подаются (1, 0) и (0, 1).

Для ускорения счета счетные триггеры пропускают запрещенные состояния, при этом логический уровень сигнала COD низкий, а сигнал OUTC изменяется с высокого уровня на низкий. Данное решение дает возможность по каждому тактовому импульсу формировать новую перестановку, обеспечивая высокую производительность генератора.

Заключение

Предложенное устройство представляет собой функциональный генератор перестановок. Матричная структура позволяет реализовать генератор перестановок чисел $(0, 1, \dots, n)$ для произвольного значения $n = 2^k$. Генератор имеет аппаратную реализацию, сложность которой растет практически линейно с ростом n , что дает возможность формирования перестановок для больших значений n . Сформулированные требования на ограничения управляющих кодов позволяют сделать работу генератора более эффек-

тивной и минимизировать пространство для их хранения. При использовании в управляющих цепях генератора синхронно-асинхронных счетчиков задержка выдачи очередного сочетания составляет около $\log_2^2(n) \cdot \tau$, где τ — задержка на счетном триггере.

Список литературы

1. Курейчик В. М., Глушань В. М., Щербаков Л. И. Комбинаторные аппаратные модели и алгоритмы в САПР // М.: Радио и связь, 1990.
2. U. S. Pat. No. 6,934,388 131 B1 H04L9/00 Method and Apparatus for Random Permutations / Clark James Monroe. Aug. 23, 2005.
3. U. S. Pat. No. 5,734,721 B1 H04L9/00 Anti-spoof without error extension (ANSWER) / Clark James Monroe. March 31, 1998.
4. Ritter T. Transposition Cipher with Pseudo-Random Shuffling: The Dynamic Transposition Combiner // Cryptologia. 1991. V. 15 (1). P. 1—17.

5. U. S. Pat. No. 7,519,795 B2 G06F 9/30 Method And System For Performing Permutations With Bit Permutation Instructions / Ruby B. Lee, Zhijie Shi. Apr. 14, 2009.
6. Молодченко Ж. А., Овчинников С. В., Сотов Л. С., Харин В. Н. Модели аппаратных акселераторов перестановок бинарных множеств // Гетеромагнитная микроэлектроника. Саратов: Изд-во Саратовского ун-та, 2008. Вып. 4. С. 11—22.
7. Portz M. A generalized description of DES-based and Benes-based permutation generators // Lecture Notes in Computer Science Springer Berlin / Heidelberg. 1993. V. 718. P. 397—409.
8. Yang X., Vachharajani M., Lee R. B. Fast Subword Permutation Instructions Based on Butterfly Networks // Proceedings of SPIE, Media Processor. 2000. January. P. 80—86.
9. Молодченко Ж. А., Сотов Л. С., Харин В. Н. Динамическое форматирование представлений объектов реляционных СУБД на основе кластерных транспозиций // Естественные и технические науки. М.: Изд-во компании "Спутник +". 2007. № 6 (32). С. 224—226.
10. Thorsten G., Stan L., Grant M., Stuart S. System Design with SystemC. Kluwer Academic Publishers, Boston/Dordrech/London, 2002.
11. Amirfathi M., Holtmann U., Ramachandran L., Toma H. and Zhang Y. CoCentric System Studio // Synthesizable SystemC RTL Code Generation, Synopsys, July 2001.

УДК 519.2:004.421

Н. Т. Абдуллаев, канд. техн. наук, зав. каф.,
Т. А. Алиев, канд. техн. наук, доц.,
О. А. Дышин, канд. физ.-мат. наук, ст. науч. сотр.
Азербайджанская государственная
нефтяная академия, г. Баку

Алгоритм вычисления статистической оценки обобщенных размерностей множеств

На основе метода нахождения всех действительных корней нелинейных уравнений предложен алгоритм построения статистической оценки обобщенных размерностей и спектра сингулярностей, используемых для характеристики фракталов и странных аттракторов.

Ключевые слова: обобщенные размерности, множества, фракталы, статистическая оценка, мультифрактальный анализ

Введение

Задача нахождения размерности множества A возникает как в теоретических, так и в чисто прикладных исследованиях. Зная размерность, можно оценить число переменных, необходимых для описания процесса. Данные переменные часто называют параметрами порядка, поскольку они определяют поведение процесса в целом. Этот

принцип, названный подчинением мод, положен в основу синергетики [1].

Для динамических систем множество A может быть гладким многообразием. Однако многие явления в физике, химии, биологии имеют фрактальную структуру [2]. Наличие фрактала в системе (т. е. хаоса) и определяется по дробности размерности [2—4].

К числу фракталов относят геометрические объекты — линии, поверхности, тела, которые имеют сильно изрезанную форму и демонстрируют некоторую повторяемость в широком диапазоне масштабов. Повторяемость может быть полной (в этом случае говорят о регулярных фракталах), либо может наблюдаться некоторый элемент случайности (такие фракталы называют случайными). Структура случайных фракталов на мелких масштабах не является идентичной структуре всего объекта, но их статистические характеристики совпадают, и свойства самоподобия сохраняются после усреднения по статистически независимым реализациям объекта [5].

Для количественного описания фракталов обычно бывает достаточно одной величины — размерности Хаусдорфа D_H или показателя скейлинга τ , описывающего сохраняемость геометрии или статистических характеристик при изменении масштаба и не зависящего от размерности q элементарных множеств, покрывающих рассматриваемое множество. Однако встречается много явлений, которые требуют распространения понятия фрактала на сложные структуры с более чем

одним показателем скейлинга (переменного показателя $\tau(q)$). Такие структуры часто характеризуются целым спектром показателей и размерность Хаусдорфа является лишь одним из них. Именно сложные фракталы, называемые мультифракталами, как правило, и встречаются в природе, тогда как простые самоподобные объекты (монофракталы) представляют собой идеализацию реальных явлений. Фактически мультифрактальный подход означает, что изучаемый объект можно разделить на части, каждая из которых обладает своими свойствами самоподобия [6].

Для мультифракталов величина D_H заменяется на величину $D(q)$, которая уже явно зависит от вещественного параметра q и называется обобщенной размерностью. Впервые это понятие было введено в работе [7] для множеств, имеющих канторову структуру. Фракталы, характеризующие пространственную упорядоченность (многочисленные примеры которых дает работа [2] и, в частности канторово множество), имеют существенное отличие от множеств, возникающих при изучении временной упорядоченности — странных аттракторов, определения которых введены в работах [8, 9]. Численное нахождение обобщенных размерностей множеств (странных аттракторов) обычно основывается на построении разбиения сетью размера ε с заданной вероятностью попадания точек множества A в каждую ячейку ε -сети и нахождения предела при $\varepsilon \rightarrow 0$ [10].

В настоящей работе для нахождения обобщенной размерности в смысле определения [8, 9] используется подход [11], основанный на вычислении средних расстояний до ближайших точек между элементами выборки точек $x_1, \dots, x_n$ исследуемого множества $A \subset R^m$. Статистическая оценка обобщенной размерности $D(q)$, полученная [11] для любых вещественных q , связана с нахождением при фиксированном n вещественных корней нелинейного уравнения. Для решения последнего нами применяется метод поиска всех действительных некротных корней нелинейного уравнения [12].

1. Процедура статистической оценки обобщенной размерности множества

Приведем определение обобщенной размерности [8, 9], основанное на понятии функции разбиения $\beta(q)$ множества с вероятностной мерой.

Пусть дано множество $A \subset R^m$ и мера μ на A , которая является регулярной борелевской [13] и вероятностной мерой, т. е. $\mu(A) = 1$. Функция разбиения $\beta(q)$ меры μ равна

$$\beta(q) = - \lim_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \log \sum_{i: \mu(K_i) > 0} \mu(K_i)^q, \quad (1)$$

где предел берется по всем разбиениям пространства R^m (фактически, окрестности A) сеткой размера ε , чтобы K_i — i -я ячейка сетки удовлетворяла условию

$$\exists x_0 \in K_i: B(x_0, \varepsilon/a) \subset K_i \subset B(x_0, \varepsilon), \mu(K_i) > 0.$$

Здесь и далее через $B(x, \varepsilon)$ обозначается открытый шар с центром в точке x и радиуса ε ($a > 1$ — некоторая константа).

Отметим, что более часто функцией разбиения (свободной энергией) меры μ называют функцию $\tau(q) = -\beta(q)$.

Если $\beta(q)$ существует, то обобщенные размерности определяются по формуле

$$D(q) = \frac{\beta(q)}{1-q}. \quad (2)$$

При $q = 1$ этому выражению также можно придать смысл, если считать его равным $-\frac{d\beta(q)}{dq} \Big|_{q=1}$.

Получаемое при этом выражение

$$D(1) = -\frac{d\beta(q)}{dq} \Big|_{q=1} \quad (3)$$

называется информационной размерностью. Так как $\beta(1) = 0$ [11], выражение (3) вытекает из формулы (2) после применения правила Лопиталя.

Второе определение обобщенных размерностей введено в работе [7]

$$D(q) = \frac{1}{q-1} \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\log \varepsilon} \log \sum_{i=1}^n \mu(B(x_i, \varepsilon))^q, \quad (4)$$

где $A \subset \bigcup_i B(x_i, \varepsilon)$ и суммирование идет по тем i , для которых $\mu(B(x_i, \varepsilon)) > 0$.

Совпадение этих двух определений для диаметрально регулярных [13] мер доказано в [14]. В этом частном случае мера Хаусдорфа d_H совпадает с емкостью множества d_c , определяемой в виде [4]:

$$d_c = \lim_{\varepsilon \rightarrow 0} \log N(\varepsilon) / \log(1/\varepsilon), \quad (5)$$

где через $N(\varepsilon)$ обозначено минимальное число шаров радиуса ε , необходимое, чтобы покрыть изучаемое множество. Размерность d_c называют также размерностью Минковского и обозначают d_M [5]. В общем же случае фрактальных множеств имеет место неравенство $d_M \geq d_H$. Компьютерные алгоритмы вычисления размерности Минковского d_M приведены подробно в работе [5].

В работе [11] предложена следующая статистическая оценка (точнее, семейство оценок, зависящих от целого параметра k) обобщенных размерностей (2) меры μ .

Пусть дана последовательность независимых случайных величин $x_1, x_2, \dots, x_n$, одинаково распределенных по мере μ с компактным носителем $A \subset \mathbb{R}^m$ ($\mu(A) = 1$). Пусть дано целое число k . Вводится вспомогательная случайная величина, зависящая от параметра β :

$$r_n(\beta) = \frac{1}{n} \sum_{j=1}^n (\min_{i \neq j}^{(k)} |x_i - x_j|)^\beta, \quad (6)$$

где $\min^{(k)}\{X_1, \dots, X_p, \dots, X_n\} = X_k$ при $X_1 \leq X_2 \leq \dots \leq X_n$ означает, что в последнем ряду чисел берется стоящее на k -й позиции число X_k .

Затем определяется случайная величина $\beta_n(q)$ как корень уравнения

$$q = 1 + \frac{\log r_n(\beta)}{\log n}. \quad (7)$$

За оценку обобщенных размерностей множества A с мерой μ (говорят также об обобщенных размерностях меры μ) принимается случайная величина

$$D_n(q) = \frac{\beta_n(q)}{1-q}. \quad (8)$$

Для самоподобных фракталов показано [11], что функция $\beta_n(q)$, определенная в уравнении (7), при $n \rightarrow \infty$ сходится к функции $\beta(q)$, определенной в выражении (1).

Оценка (8) обобщенных размерностей $D(q)$, определенных по формуле (2), для фрактального множества A из $\mathbb{R}^m$ получена в работе [11] на основе следующих рассуждений.

Пусть задача — выборка $x^0 = \{x_1, x_p, x_n\}$ точек множества A . Для каждой точки $x_i \in x$ вычислим

$$\rho_{n,i}^{(k)} = \min_{j:j \neq i}^{(k)} |x_j - x_i|, \quad (9)$$

где $|x|$ — длина вектора x в евклидовом пространстве $\mathbb{R}^m$.

Рассмотрим набор шаров $B(x_i, \rho_{n,i}^{(k)})$ с центром в точке x_i и радиуса $\rho_{n,i}^{(k)}$. Предположим, что при достаточно большом n эти шары образуют покрытие множества A , и вычислим сумму

$$S_n(\beta) = \sum_{i=1}^n (\rho_{n,i}^{(k)})^\beta, \quad (\beta > 0). \quad (10)$$

Выдвигается гипотеза, что при $n \gg 1$ значение случайной величины $S_n(\beta)$ удовлетворяет асимптотическому равенству

$$S_n(\beta) = n^{q+0(1)}, \quad (11)$$

где $0(1)$ — слагаемое, стремящееся к нулю при $n \rightarrow \infty$. Отметим, что равенство (11) эквивалентно предельному соотношению

$$q = \lim_{n \rightarrow \infty} \frac{\log S_n(\beta)}{\log n}. \quad (12)$$

Предположение (11) не противоречит идее определения размерности Хаусдорфа [4]: при достаточно большом положительном значении β в пределе сумма Хаусдорфа равна нулю, а при малом β — бесконечности. Инфимум тех значений β , при которых эта сумма равно нулю (т. е. $S_n(\beta) = 0$) и есть размерность Хаусдорфа множества A . В этой связи за оценку размерности Хаусдорфа можно принять корень уравнения $S_n(\beta) = 1$, т. е. в соотношении (12) $q = 0$.

Производные по β функции $S_n(\beta)$ удовлетворяют неравенствам

$$\begin{aligned} S'_n(\beta) &= \sum_{i=1}^n (\rho_{n,i}^{(k)})^\beta \ln \rho_{n,i}^{(k)} < 0, \\ S''_n(\beta) &= \sum_{i=1}^n (\rho_{n,i}^{(k)})^\beta \ln^2 \rho_{n,i}^{(k)} > 0 \end{aligned} \quad (13)$$

при $\rho_{n,i}^{(k)} < 1$ (что выполняется при большом n). Поэтому из уравнения, полученного из (11) отбрасыванием членов $0(1)$:

$$S_n(\beta) - n^q = 0, \quad (14)$$

можно найти корень $\beta_n(q)$. Ясно, что $\beta_n(1) = 0$ и $\beta_n(q) \rightarrow \infty$ при $q \rightarrow \infty$.

За оценку обобщенных размерностей принимается функция (8), в которой $\beta_n(q)$ — корень уравнения (14). От суммы $S_n(\beta)$ легко перейти к ее среднему $r_n(\beta)$, определенному в формуле (5). В соответствии с предположением (11) имеем

$$r_n(\beta) = n^{q-1+0(1)}. \quad (15)$$

Логарифмируя последнее равенство, получим уравнение (6), которое было предложено выше для определения функции $\beta_n(q)$.

Несмотря на то, что вывод статистической оценки (8) базируется на определении размерности Хаусдорфа и опирается на предположение $\beta > 0$, формально сумма $S_n(\beta)$ (10) определена и для $\beta < 0$. Если при этом асимптотическое равенство (11) имеет место, то ясно, что в нем $q > 1$ и оценка размерности будет определена для $q > 1$. Как показано в работе [11], для гладких многообразий и самоподобных фракталов равенство (11) становится справедливым при $\beta > \beta^{(k)}$, где $\beta^{(k)} \rightarrow \infty$ при $k \rightarrow \infty$. Это позволяет находить обобщенные размерности и при отрицательных значениях β (соответственно при $q > 1$), но в этом случае k

должно удовлетворять некоторому ограничению (см. ниже условие 4⁰).

Для указанных двух частных случаев компактного множества $A \subset \mathbb{R}^m$ с борелевской вероятностной мерой μ случайные величины $D_n(q)$ сходятся по вероятности при $n \rightarrow \infty$ к обобщенным размерностям $D_n(q)$, определенным в формуле (2). Таким образом, оценка размерности по формулам (6)–(8) является состоятельной и не зависит от распределения случайных величин x_i .

Эта оценка определяется по последовательности $\{\beta_n(q)\}$ корней уравнения (14) или эквивалентного уравнения, получаемого из равенства (15) отбрасыванием бесконечно малой $O(1)$:

$$r_n(\beta) - n^q = 0. \quad (16)$$

Остановимся отдельно на выборе числа k в выражениях (6) и (9). Согласно теореме 1 [11] статистические характеристики $E(\rho_{n,i}^{(k)})^\beta$ и $D(\rho_{n,i}^{(k)})^\beta$ (E и D — знаки математического ожидания и дисперсии) существуют, если выполняются следующие условия:

1⁰. A — компакт в $\mathbb{R}^m$ с борелевской вероятностной мерой μ .

2⁰. Случайные величины $x_1, x_2, \dots, x_n$ независимы и одинаково распределены на A по этой мере.

3⁰. Функция распределения $F(t, y)$ величины $|x_i - y|$ непрерывна и удовлетворяет ограничениям

$$C_1 t^{\alpha_1} < F(t, y) < C_2 t^{\alpha_2}, \quad \forall t > 0, \quad \forall y \in A, \quad (17)$$

где $C_1, C_2, \alpha_1, \alpha_2$ — некоторые константы.

4⁰. Величины k и β , применяющиеся в формулах (6) и (10), не зависят от n и удовлетворяют ограничению

$$(k - 2) - \alpha_2 + 2\beta > 0. \quad (18)$$

Пусть, кроме этого, выполняется условие 5⁰.

5⁰. Существует

$$\lim_{n \rightarrow \infty} \frac{\log E(\rho_{n,i}^{(k)})^\beta}{\log n} = q(\beta) - 1, \quad (19)$$

где $q(\beta)$ не зависит от k и удовлетворяет неравенству

$$q(2\beta) < 2q(\beta), \quad \forall \beta. \quad (20)$$

Тогда по теореме 1 [11] верно предельное соотношение

$$D(r_k(\beta)n^{1-q(\beta)}) \rightarrow 0, \quad n \rightarrow \infty, \quad (21)$$

где случайные величины $\rho_{n,i}^{(k)}$ определены в формуле (9), а $r_n(\beta)$ в (6).

В работе [11] показано, что ограничения (18)–(20) всегда выполняются в случае гладких многообразий и в случае самоподобных фракталов с самоподобными мерами.

Найдем достаточное условие, налагаемое на параметр k при фиксированном β , обеспечивающее выполнение ограничения (18) в общем случае множества A , удовлетворяющее условию 1⁰.

Пусть при фиксированном $\varepsilon_0 \geq 1$ шары $B(x_i, \varepsilon_0)$ ($i = 1, \dots, n$) покрывают множество A и $B'(x_i, \varepsilon_0)$ — пересечение шара $B(x_i, \varepsilon_0)$ с множеством A . Тогда $A = \cup B'(x_i, \varepsilon_0)$ и любое y из A будет принадлежать $B(x_i, \varepsilon_0)$ с вероятностью $p_i(y) \geq 0$, при этом $\sum_i p_i(y) \geq 1$, так как шары $B(x_i, \varepsilon_0)$ (и, стало быть, множества $B'(x_i, \varepsilon_0)$) могут перекрываться. Для $\forall y \in A, t > 0$ и $\forall i \in I = \{1, \dots, n\}$ функция распределения $F(t, y) = P\{|x_i - y| < t, i \in I\}$ принимает значение равное 1 при $t > \varepsilon_0$ и меньшее, чем

$$p_i(y) \left(\frac{t}{\varepsilon_0}\right)^m, \quad \text{при } t \leq \varepsilon_0, \quad \text{поскольку } \left(\frac{t}{\varepsilon_0}\right)^m \text{ равно отношению объемов шаров } B(x_i, t) \text{ и } B(x_i, \varepsilon_0) \text{ в пространстве } \mathbb{R}^m. \text{ Следовательно, } 0 < F(t, y) < \frac{1}{\varepsilon_1^m} t^m,$$

где ε_1 — число, близкое к 1, но меньшее 1. Таким образом, ограничение (17) выполняется при $\alpha_1 = 0, \alpha_2 = m, C_1 = 0, C_2 = 1/\varepsilon_1^m$.

Поэтому ограничение (18) будет удовлетворяться при условии 6⁰:

6⁰. Для $\beta > 0, k \geq 2$; для $\beta = 0, k \geq 3$;

$$\text{для } -\alpha < \beta < 0 \ (\alpha > 0) \ k \geq 2 \left(1 + \left\lceil \frac{a}{m} \right\rceil\right) + 1, \quad (22)$$

где $[a]$ — целая часть числа a .

Как показано в работе [11], если A — гладкое компактное многообразие в $\mathbb{R}^m$, имеющее топологическую размерность s ($s \leq m$), то в ограничении (17) $\alpha_1 = \alpha_2 = s$ и $q(\beta) = 1 - \beta/s$ ($q(\beta)$ — функция, обратная к $\beta(q)$). Из последнего равенства и ограничения (18) вытекает условие 7⁰.

7⁰.

$$k > 2q. \quad (23)$$

2. Алгоритм численной реализации статистической оценки обобщенной размерности

Уравнение (14), как и эквивалентное ему уравнение (16), при фиксированных q и n представляет собой нелинейное уравнение относительно β .

Одним из наиболее трудных вопросов при решении нелинейных уравнений и систем уравнений является проблема нахождения всех корней, т. е. задача выделения подмножеств, содержащих один корень. Итерационные методы решения систем нелинейных уравнений [13, 14] работают в условиях изолированности решения и наличия "хорошего" (близкого к решению) начального

приближения в области, в которой оператор, порождающий итерационный процесс, будет сжимающим.

Применяемые на практике подходы к решению этой проблемы (использование априорной информации о положении корней, "сеточные" методы) не могут гарантировать нахождения всех корней нелинейного уравнения. Как было отмечено в предыдущем разделе, функция $S_n(\beta)$ в уравнении (14) дважды дифференцируема по β . Поэтому эффективным методом поиска корней уравнения (14) будет предложенный в работе [12] метод нахождения всех действительных некротных корней нелинейного уравнения

$$f(x) = 0 \quad (24)$$

на отрезке $[c, d]$, где $f(x)$ — дважды дифференцируемая функция. В работе [16] данный метод обобщается на системы нелинейных уравнений.

Идея метода, изложенного в работе [12], состоит в разбиении исходного множества, в котором ищутся корни, на совокупности подмножеств, в каждом из которых либо отсутствуют корни, либо $f'(x) \neq 0$, и уравнение в этом подмножестве имеет не более одного корня. В последнем случае в подмножестве либо отсутствуют корни, либо содержится единственный корень.

В качестве уравнения (24) будем рассматривать уравнение (14) для $\beta \in [c, d] \subset \mathbb{R}^m$ при фиксированных n и q , т. е. $x = \beta$, $f(x) = S_n(\beta) - n^q$ и $x \in D_0 = [c, d]$. В этом случае в силу неравенств (13) функция $f(x)$ дважды дифференцируема в любой области $D_0 \subset R$ и уравнение (24) либо не имеет корней, либо имеет единственный корень в D_0 .

Будем полагать, что объем n выборки $x = (x_1, \dots, x_n)$ достаточно большой, и $n > k$, причем $x_1, \dots, x_n$ так располагаются в A , что $\rho_{n,i}^{(k)} < 1$, $\forall i = 1, \dots, n$, где k удовлетворяет условиям 6⁰, 7⁰ для $\beta \in [-a, a] \supseteq [c, d]$. При такой выборке x функция $f(x)$ для всех $x \in D_0 = [c, d]$ удовлетворяет ограничениям

$$|f'(x)| \leq M_1, \quad |f''(x)| \leq M_2, \quad (25)$$

где

$$M_1 = |\ln \min_i \rho_{n,i}^{(k)}|, \quad M_2 = |\ln^2 \min_i \rho_{n,i}^{(k)}|. \quad (26)$$

Так как $f'(x) < 0$, то $f(x)$ и $f'(x)$ не имеют общих нулей, поскольку $f'(x) < 0$ для $x \in D_0$.

Пусть $x_0 = 1/2(c + d)$. Если $f(x)$ удовлетворяет условию

$$|f(x_0)| > \frac{1}{2} M_1(d - c), \quad (27)$$

то уравнение (24) не имеет корней в D_0 (теорема 1 из работы [12]), а если $f(x)$ удовлетворяет условию

$$|f'(x_0)| > \frac{1}{2} M_2(d - c), \quad (28)$$

то уравнение (24) имеет не более одного корня в D_0 (теорема 3 из [12]), т. е. уравнение (24) может либо не иметь корней в D_0 , либо иметь единственный корень. Поскольку $f'(x) < 0 \forall x \in D_0$, то наличие корня в D_0 устанавливается проверкой неравенства

$$f(c) \cdot f(d) < 0. \quad (29)$$

Если неравенство (29) выполняется, то уравнение (24) имеет в D_0 единственный корень. В противном случае корней в D_0 нет.

При хорошем начальном приближении корень уравнения (24) определяется методом Ньютона:

$$x^{v+1} = x^v - \frac{f(x^v)}{f'(x^v)}. \quad (30)$$

Как известно [16], скорость сходимости метода Ньютона оценивается следующим образом. Если при $0 < a, a_1, a_2 < \infty$ выполняются условия

$$|f'(x)| \leq 1/a_1, \quad x \in \Omega_a = \{x: |x - x^*| < a\}; \quad (31)$$

$$|f(x_1) - f(x_2) - f'(x_2)(x_1 - x_2)| \leq a_2|x_2 - x_1| \quad \text{при } x_1, x_2 \in \Omega_0, \quad (32)$$

то итерационный процесс (30) сходится с квадратичной скоростью

$$|x^{v+1} - x^*| \leq \lambda |x^v - x^*|^2 \quad (33)$$

и оценкой погрешности $|x^v - x^*| \leq \varepsilon_v(x^*)$ с

$$\varepsilon_v(x^*) = \lambda^{-1}(\lambda |x^0 - x^*|)^{2^v}, \quad (34)$$

где $\lambda = a_1 a_2$; x^0 и x^* — соответственно начальное приближение и корень уравнения (24), при этом полагаем, что $D_0 \subseteq \Omega_a$ и, следовательно, (31) удовлетворяется с $a_1 = 1/M_1$. Из формулы Тейлора с остаточным членом в форме Лагранжа следует, что неравенство (32) выполняется с $a_2 = aM_2/a$, где M_2 — мажоранта для $f''(x)$ в D_0 .

На основании теоремы 4 работы [12], существует такое $\Delta l > 0$, что из условия $d - c < \Delta l$ следует выполнение одного из неравенств (27), (28).

Из изложенных выше рассуждений приходим к следующему алгоритму вычисления обобщенных размерностей.

Назначение алгоритма: вычислить для заданного $q(q \neq 1)$ значение обобщенной размерности $D(q)$ с погрешностью ε , если $D(q)$ существует и конечен.

Входные данные: $\varepsilon, q, a, m, n_1, n$; координаты точек $x_1, \dots, x_n$ множества A из $\mathbb{R}^m$, для которого нужно вычислить $D(q)$; формулы (22), (23) для вычисления значения k ; формулы (6) для вычисления $r_n(\beta)$; формулы (9), (10), (14) для вычисления функции $f(x) = S_n(\beta)|_{\beta=x} - n^q$.

Выходные данные: оценка с погрешностью ε обобщенной размерности $D(q)$ при фиксированном q .

Алгоритм состоит из следующих пунктов:

1. Зададим ε, m, q, a ($a > q - 1$) и P_0 .
2. Зададим величину k , удовлетворяющую ограничениям (22) и (23).
3. Зададим $n_1 = \max\{4k; 100\}$, $n = n_1/2$. Положим $\lambda = 0$.
4. Зададим координаты точек $x_1, \dots, x_{n_1}$ из множества A , которые удовлетворяют неравенству

$$\rho_{n,i}^{(k)} < 1, \quad \forall i = 1, \dots, n_1,$$

для этого достаточно, чтобы в выборке $\{x_1, \dots, x_{n_1}\}$ имелись k пар точек (x_i, x_j) ($i \neq j$), расстояние $|x_i - x_j|$ между которыми было бы меньше единицы.

5. Опишем функции

$$f(x) = S_n(\beta)|_{\beta=x} - n^q \text{ и } f'(x) = \sum_{i=1}^n (\rho_{n,i}^{(k)})^\beta \ln \rho_{n,i}^{(k)} |_{\beta=x}.$$

6. Вычислим константы

$$M_1 = |\ln \min_{i=1, n} \rho_{n,i}^{(k)}|, \quad M_2 = M_1^2.$$

7. Полагаем $\mu = 0, c = -a; d = a$.
8. Проверяем неравенство (27). Если неравенство выполнено, то переходим к п. 9, в противном случае — к п. 1.
9. Проверяем условие $(c = -a) \wedge (d = a)$. Если условие выполнено, то полагаем $a := 2a$ и переходим к п. 2, в противном случае — к п. 11.
10. Проверяем неравенство (28). Если неравенство выполнено, то переходим к п. 12, в противном случае — к п. 13.
11. Если $\mu = 1$, то переходим к п. 17, в противном случае — к п. 15.
12. Проверяем неравенство (29). Если оно выполнено, то переходим к п. 14, в противном случае — к п. 13.
13. Если $\mu = 0$, то переходим к п. 15, в противном случае — к п. 17.
14. Проверяем неравенство

$$d - c < \frac{4}{a} \cdot \frac{M_1}{M_2}.$$

Если оно выполнено, то переходим к п. 18, в противном случае полагаем $\mu = 0$ и выполняем п. 14.

15. Вычисляем $c_1 = c, d_1 = (c + d)/2, c_2 = d_1, d_2 = d$.

16. Полагаем $\mu := 1, c := c_1, d := d_1$ и идем на п. 8.

17. Полагаем $\mu := 0, c := c_2, d := d_2$ и идем на п. 8.

18. Полагаем $v = 0$ и вычисляем $x^v = (c + d)/2$.

19. Вычисляем x^{v+1} по формуле (30).

20. Вычисляем

$$\varepsilon_{v+1} = \frac{2}{a} \cdot \frac{M_1}{M_2} \left(\frac{a(d-c)}{4} \cdot \frac{M_1}{M_2} \right)^{2^{v+1}}.$$

21. Проверяем неравенство $\varepsilon_{v+1} \leq \varepsilon|1 - q|$. Если оно выполнено, то переходим к п. 22, в противном случае полагаем $x^v := x^{v+1}, v := v + 1$ и переходим к п. 19.

22. Положим $\beta_n := x^{v+1}$. Вычислим $r_n(\beta_n)$ и $r_n(2\beta_n)$ по формуле (6) и величины

$$Q_n(\beta_n) = 1 + \frac{\log r_n(\beta_n)}{\log n}, \quad Q_n(2\beta_n) = 1 + \frac{\log r_n(2\beta_n)}{\log n},$$

$$c_n = \frac{1}{n} \sum_{i=1}^n (\rho_{n,i}^{(k)} - r_n(\beta_n))^2, \quad P = 1 - \frac{c_n}{n\varepsilon^2},$$

где c_n — выборочная дисперсия случайных величин $\rho_{n,1}^{(k)}, \dots, \rho_{n,n}^{(k)}$; $1 - c_n/(n\varepsilon^2)$ — правая часть неравенства Чебышева.

23. Если $\lambda = 0$, то полагаем $\lambda := \lambda + 1, n := n + 1$ и переходим на п. 5.

24. Проверяем неравенство $|q - Q_n(\beta_n)| < \varepsilon$. Если оно выполнено, то идем на п. 25, в противном случае — на п. 28.

25. Проверяем неравенство $|q - Q_n(2\beta_n)| < \varepsilon$. Если оно выполнено, то идем на п. 26, в противном случае — на п. 28.

26. Проверяем неравенство $Q_n(2\beta_n) < 2Q_n(\beta_n)$. Если оно выполнено, то идем на п. 27, в противном случае — на п. 28.

27. Проверяем неравенство $P \geq P_0$. Если оно выполнено, то идем на п. 29, в противном случае — на п. 28.

28. Если $n < n_1$, то полагаем $n := n + 1$ и переходим на п. 5, в противном случае идем на п. 30.

29. Вычислим $D(q) = \beta_n/(1 - q)$. Печатаем $D(q)$ и P .

30. Печатаем: используемая оценка обобщенной размерности несостоятельна при заданном значении q .

3. Мультифрактальный анализ процессов на основе обобщенных размерностей их геометрической структуры

Фрактальные подходы к анализу процессов весьма разнообразны в том смысле, что фракталь-

ному анализу могут подвергаться различные, связанные с процессом, геометрические объекты. Как фрактал в ряде случаев можно рассматривать график связанной с процессом величины. Фрактальным может быть, например, множество точек пересечения величиной определенного уровня. Квадрат амплитуды процесса можно рассматривать как плотность точек на временной оси, имеющую фрактальную структуру. Вообще говоря, плотности различных описывающих процесс величин могут иметь фрактальное распределение во времени и пространстве.

Один из наиболее содержательных на сегодня подходов к фрактальному анализу процессов опирается на предположении о том, что процесс порождается динамической системой конечной размерности с хаотическим поведением. Траектория такой системы в фазовом пространстве представляет собой фрактальное множество, называемое странным аттрактором. Анализ процесса, порожденного такой системой, основан на применении алгоритма Такенса [17], позволяющего восстановить траекторию системы в фазовом пространстве по временной зависимости одной связанной с системой величины. Анализ процессов, порожденных конечномерной хаотической динамикой, его теоретические и практические стороны подробно описаны в литературе (см. например [18, 19]).

Обратимся к идее мультифрактального разложения [11]. Для каждой точки x множества A вводится точечная размерность меры μ :

$$d_v(x) = \lim_{\varepsilon \rightarrow 0} \frac{\log \mu(B(x, \varepsilon))}{\log 2q}. \quad (35)$$

Используя точечную размерность, множество A можно разбить на подмножества $A = \cup A_\alpha$, где $A_\alpha = \{x \in A: d_\mu(x) = \alpha\}$. Далее для каждого множества A_α находим

$$f(\alpha) = \dim_H A_\alpha, \quad (36)$$

где $\dim_H$ — размерность Хаусдорфа.

Функция

$$\beta^*(\alpha) = \inf_{q \in \mathbb{R}} (q\alpha + \beta(q)) \quad (37)$$

называется преобразованием Лежандра функции $\beta(q)$. Если существует производная $\beta'(q)$ и функция $\beta(q)$ выпукла, то

$$\beta^*(\alpha) = q\alpha + \beta(q), \quad \alpha = -\beta'(q). \quad (38)$$

В общем случае доказано (см., например [20]), что

$$f(\alpha) \leq \beta^*(\alpha). \quad (39)$$

Равенство

$$f(\alpha) = \beta^*(\alpha) \quad (40)$$

называется мультифрактальным формализмом. Описание ряда множеств и мер, для которых найдена функция $f(\alpha)$ и доказано равенство (40), приведено в работе [11]. К ним относятся: 1) самоподобные фракталы с самоподобной мерой; 2) более общий класс фракталов — *digrah recursive fractals*; 3) случайные самоподобные фракталы; 4) канторовские множества на прямой, построенные нелинейными гладкими сжатиями (*cookie-cutter sets*).

Пусть для фрактального множества A выполнено равенство (40). Тогда введя функцию

$$\tau(q) = -\beta(q) = (q - 1)D(q), \quad (41)$$

из (34) и (36) получим

$$f(\alpha) = q \cdot \alpha - \tau(q), \quad \alpha(q) = \frac{d\tau(q)}{dq}. \quad (42)$$

Таким образом, зная зависимость $\alpha(q)$, называемую α -спектром, можно вычислить обобщенную размерность $D(q)$, определенную формулой (2) и наоборот.

Пусть $d_{j,k}$ — вейвлет-коэффициенты сигнала $g(x)$, $x \in \mathbb{R}$, полученные применением к функции $g(x)$ дискретного вейвлет-преобразования (ДВП) на уровне разрешения j [21].

Построим частичные функции

$$Z_q(j) = \sum_k (\sup_{j' \leq j} |d_{j',k}|)^q. \quad (43)$$

Было показано [22, 23], что в случае фрактального сигнала эта сумма должна вести себя при больших j как

$$Z_q(j) \sim 2^{j[\tau(q) + q/2]}, \quad (44)$$

т. е.

$$\log Z_q(j) \sim 2j[\tau(q) + q/2]. \quad (45)$$

Итак, необходимым условием того, что сигнал $g(x)$ обладает свойствами самоподобия, является линейная зависимость $\log Z_q(j)$ от номера уровня j . Если это требование удовлетворено, то зависимость функции $\tau(q)$ от ранга момента q определяет, является ли данный сигнал монофрактальным или мультифрактальным. Для монофрактального сигнала $D(q) = \text{const}$ при различных q и $\tau(q)$ — прямая линия на плоскости (q, τ) , а для мультифрактального сигнала $D(q)$ — убывающая функция от q и $\tau(q)$ — возрастающая функция от q . Монофрактальные сигналы характеризуются одной размерностью, тогда как для описания мультифрактальных сигналов необходим целый набор таких размерностей.

Дифференцируя $f(\alpha)|_{\alpha=\alpha(q)}$ по q , из выражения (41) находим

$$\begin{aligned} \frac{df(\alpha(q))}{dq} &= f'(\alpha)\alpha'(q) = \\ &= \left(\frac{dq}{d\alpha} \alpha + q - \frac{d\tau}{dq} \cdot \frac{dq}{d\alpha} \right) \frac{d\alpha}{dq} = \\ &= \alpha + q\alpha'(q) - \alpha = q\alpha'(q), \end{aligned}$$

откуда

$$\left. \frac{df}{d\alpha} \right|_{\alpha=\alpha(q)} = q. \quad (46)$$

Если μ — малое положительное число, то

$$\begin{aligned} \frac{d}{dq} [f(\alpha) - q(\alpha - \mu)] &= -\alpha + \mu; \\ \frac{d}{dq} [f(\alpha) - q(\alpha + \mu)] &= -\alpha - \mu, \end{aligned}$$

т. е. $f'(\alpha)$ при переходе через стационарную точку $\alpha = \alpha(q)$ ($f'(\alpha)|_{\alpha=\alpha(q)} = 0$) меняет знак с (+) на (−). Следовательно, в этой точке $f(\alpha)$ принимает максимальное значение и, следовательно,

$$\left. \frac{d^2 f}{d\alpha^2} \right|_{\alpha=\alpha(q)} < 0. \quad (47)$$

Из формул (41), (42) следует, что $D(0) = f(\alpha(0))$, и в соответствии с формулами (46), (47) функция $f'(\alpha)$ имеет колокообразную форму на плоскости (α, f) и ордината максимума графика $f(\alpha)$ характеризует Хаусдорфову размерность фрактала, соответствующего сигналу $g(x)$.

Из (45) при достаточно больших j находим

$$\tau(q) \sim \frac{1}{j} \log Z_q(1) - \frac{q}{2}, \quad (48)$$

и затем определяем по формуле (41) оценку для обобщенной размерности $D(q)$:

$$D(q) \sim (q - 1)^{-1} \tau(q). \quad (49)$$

Если в определении точечной размерности (35) в знаменателе дроби заменить 2ε на ε , придем ко второму определению точечной размерности (см. [4]):

$$q_p(x) = \lim_{\varepsilon \rightarrow 0} \frac{\log \mu(B(x, \varepsilon))}{\log \varepsilon}. \quad (50)$$

Разбивая теперь множество A на подмножества $A = \bigcup_{\alpha} A_{\alpha}$, где $A_{\alpha} = \{x \in A: d_p(x) = \alpha\}$, можно ввести спектральную функцию

$$\hat{f}(\alpha) = \dim_M A_{\alpha}, \quad (51)$$

где $\dim_M$ — размерность Минковского. Функцию $\hat{f}(\alpha)$ называют спектром сингулярностей [24]. Смысл этой функции состоит в следующем.

Предположим, что на множестве A задано распределение меры μ . Если это множество покрывать шарами диаметром ε , то мера шара с центром в точке x зависит от ε по степенному закону вида

$$\mu_x(\varepsilon) \sim \varepsilon^{\alpha(x)}, \quad (52)$$

где $\alpha(x)$ называется экспонентой сингулярности. Зависимость (52) удобнее переписать в виде

$$\alpha(x) = \lim_{\varepsilon \rightarrow 0} \frac{\ln \mu_x(\varepsilon)}{\ln \varepsilon}. \quad (53)$$

Чем меньше $\alpha(x)$, тем более сингулярным является распределение меры в этой точке. При равномерном распределении $\alpha(x) = 1$. Предел $\alpha(x) = 0$ соответствует распределению меры, подобному функции Дирака, и означает, что мера сосредоточена в одной точке [23].

На практике вычислить функцию $\hat{f}(\alpha)$ на основе формулы (51) весьма проблематично ввиду очень медленной сходимости при $\varepsilon \rightarrow 0$. Кроме того, значения оцениваемых характеристик могут заметно различаться для разных выбранных точек. Поэтому в теории мультифракталов предпочитают использовать специальный подход, основанный на расчете обобщенных фрактальных размерностей как глобальных характеристик, по которым можно вычислить спектр сингулярностей $\hat{f}(\alpha)$.

В рамках данного подхода в работах [22, 23] вводятся в рассмотрение так называемые частичные функции (или обобщенные частичные функции)

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} \mu_i^q(\varepsilon), \quad (54)$$

где $N(\varepsilon)$ — число элементов покрытия размером ε ; μ_i — мера i -го элемента покрытия; $q \in R$. Зависимости функций Z от ε , как правило, носят степенной характер вида [6]

$$Z(q, \varepsilon) \sim \varepsilon^{(q-1)\hat{D}(q)}, \quad (55)$$

где величина $\hat{D}(q)$ представляет собой обобщенную размерность в смысле определения (4). Коэффициент $q - 1$ включен в показатель степени в (55) для автоматического выполнения равенства $Z(1, \varepsilon) = 1$, означающего условие нормировки для меры μ . Обычно используется следующее обозначение:

$$\hat{\tau}(q) = (q - 1)\hat{D}(q), \quad (56)$$

величины $\hat{\tau}(q)$ называются скейлинговыми (характеризующими масштабное подобие) экспонентами. Мы будем обозначать их через $\hat{\tau}(q)$ в отличие от скейлинговых экспонент $\tau(q)$, определенных формулой (41). Как видно из формул (54),

(55), при $q = 0$ имеем $\lim_{\varepsilon \rightarrow 0} Z(0, \varepsilon) = d_c$, где d_c — емкость множества, определенная формулой (5).

С помощью вейвлет-преобразования можно получить оценку и для обобщенной размерности, определяемой формулой (4), которую в дальнейшем будем обозначать $\hat{D}(q)$, чтобы отличать ее от обобщенной размерности $D(q)$, определяемой формулой (2).

Для распространения концепции мультифракталов на случай функциональных зависимостей (сигналов) в работах [22, 23] был предложен метод "максимумов модулей вейвлет-преобразований" (ММВП). Суть данного метода состоит в построении частичных сумм по формуле

$$Z(q, \varepsilon) = \sum_{l \in L(\alpha)} W(a, x_l(\alpha)), \quad (57)$$

где $L(\alpha)$ — множество всех линий (l) локальных максимумов модулей вейвлет-коэффициентов, существующих на масштабе α ; $x_l(\alpha)$ характеризует положение на этом масштабе максимума, относящегося к линии l . Здесь

$$W(a, x_0) = \frac{1}{a} \int_{-\infty}^{\infty} g(x) \psi\left(\frac{x-x_0}{a}\right) dx \quad (58)$$

отличается от обычного непрерывного вейвлет-преобразования функции $g(x)$ множителем $\frac{1}{\sqrt{a}}$, который введен в работе [22] для упрощения анализа. В этом случае при $\alpha \rightarrow 0$ наблюдается простая степенная зависимость

$$W(a, x_0) \sim a^{\alpha(x_0)}. \quad (59)$$

Таким образом, если функция $g(x)$ имеет особенность в точке $x = x_0$, то ее локальное сингулярное поведение характеризуется степенным законом вида (56).

Определение (54) не подходит для отрицательных значений q , поскольку возможна ситуация, в которой $W(\alpha, < x(\alpha)) = 0$. В связи с этим на практике используется другая формула

$$Z(q, a) = \sum_{l \in L(\alpha)} \left(\sup_{a' \leq a} |W(a', x_l(a'))| \right)^q, \quad (60)$$

т. е. выбирается максимальное значение модуля коэффициентов вейвлет-преобразования вдоль каждой линии на масштабах, меньших заданного значения α . Согласно [22, 23] при достаточно малых α выполняется следующая зависимость:

$$Z(q, a) \sim a^{\hat{\tau}(q)}, \quad (61)$$

где $\hat{\tau}(q)$ определена формулой (56).

Отсюда можно вычислить

$$\hat{\tau}(q) \sim \ln Z(q, a) / \ln a \quad (62)$$

и оценку обобщенной размерности, определенной формулой (4)

$$\hat{D}(q) \sim (q-1)^{-1} \hat{\tau}(q). \quad (63)$$

Для монофрактальных объектов $\hat{\tau}(q)$ линейно зависит от q и $\alpha = \frac{d\hat{\tau}}{dq} = \text{const}$, в случае мультифракталов получается нелинейная зависимость $\hat{\tau}(q) = \alpha q - \hat{D}(q)$ с большим числом гельдеровских экспонент $\alpha(q) = \frac{d\hat{\tau}}{dq} \neq \text{const}$.

Как указано в работе [6], метод ММВП не применим для изучения сингулярностей функций, которые носят осциллирующий характер (например, в случае наличия специфических частотно-модулированных закономерностей, наблюдаемых для так называемых "чирнов" [21]). Еще одним из ограничений метода ММВП является то, что он позволяет оценить лишь верхнюю огибающую мультифрактального спектра. Это обстоятельство может приводить к ошибочной интерпретации результатов численного анализа. Во-первых, если истинный спектр $\tilde{D}(\alpha) = \hat{D}(q(\alpha))$ ($q(\alpha)$ — обратная функция к $\alpha(q)$) является дискретным, т. е. α принимает только дискретное множество значений, то огибающая будет включать "ложные" точки, представляющие собой интерполяцию дискретного спектра. Во-вторых, огибающая не позволяет идентифицировать "внутренние" точки (если они существуют), не принадлежащие верхней огибающей спектра $\tilde{D}(h)$.

Недостатков, присущих ММВП, можно избежать, используя статистическую оценку обобщенной размерности $D(q)$, описанную в разделах 1 и 2.

Заключение

1. Традиционные методы численного нахождения обобщенных размерностей фрактальных множеств основываются на построении ε -покрытия и нахождении предела при $\varepsilon \rightarrow 0$. Предлагаемый в работе [11] подход, связанный с вычислением средних расстояний до ближайших точек между элементами выборки точек $A \subset \mathbb{R}^m$, позволяет убрать эти дополнительные вычисления и существенно их упростить. Получаемая при этом оценка обобщенной размерности множества A является состоятельной для гладких многообразий с лебеговыми мерами и для самоподобных фракталов с самоподобными мерами.

2. Статистическая оценка (8) обобщенной размерности $D(q)$, определенной формулой (2), связана с нахождением при фиксированных n и q вещественных корней $\beta_n(q)$ уравнения (14). В силу

свойств (13) функций $S_n(\beta)$ уравнение (14) может иметь не более одного корня.

3. Предлагаемый в данной работе алгоритм численной реализации статистической оценки обобщенной размерности для компактного множества A из $\mathbb{R}^m$ с борелевской вероятностной мерой μ использует метод [12] поиска действительных некротных корней нелинейного уравнения $f(x) = 0$, определяемого дважды дифференцируемой функцией $f(x)$. При этом величины k и β , применяющиеся в формулах (6) и (10), при условиях (22) и (23) не зависят от n и удовлетворяют ограничениям (18), обеспечивающим вместе с условиями (19) и (20) сходимость по вероятности случайных величин $D_n(q)$ к $D(q)$ при $n \rightarrow \infty$, если для данного q существует $D(q)$.

4. Как фрактал в ряде случаев можно рассматривать график связанной с процессом величины. Квадрат амплитуды процесса можно рассматривать как плотность точек на временной оси, имеющую фрактальную структуру. С помощью дискретного вейвлет-преобразования сигнала $g(x)$ при достаточно больших уровнях разрешения j вычисляется оценка (49) обобщенной размерности $D(q)$.

5. Для распространения концепции мультифракталов на случай функциональных зависимостей (сигналов) $g(x)$ в работах [22, 23] был предложен метод "максимумов" модулей вейвлет-преобразований" (ММВП). С помощью непрерывного вейвлет-преобразования функции $g(x)$ при достаточно малых α получается оценка (63) обобщенной размерности (4) для фрактального множества, соответствующего сигналу $g(x)$.

6. Метод ММВП позволяет оценивать лишь верхнюю огибающую мультифрактального спектра и к тому же неприменим для изучения переменн-осциллирующих сигналов. Статистическая оценка (8) обобщенной размерности $D(q)$ не имеет ограничений, присущих методу ММВП, и может применяться как в случае сплошного спектра размерностей $\tilde{D}(\alpha) = D(q(\alpha))$, ($q(\alpha)$ — обратная функция к $\alpha(q)$), так и в случае дискретного спектра размерностей $\tilde{D}(\alpha)$, соответствующего дискретному набору гильбертовских экспонент α сигнала $g(x)$.

Список литературы

1. **Хакен Г.** Синергетика. М.: Мир, 1980.
2. **Mandelbrot B. B.** The fractal geometry of nature — San Francisco: Freeman, 1982.
3. **Falkoner K. J.** The geometry of fractal sets. Cambridge: Cambridge Univ. Press, 1985.
4. **Ахромеев А. В., Курдюмов С. П., Малинецкий Г. Г., Самарский А. А.** Нестационарные процессы и диффузионный хаос. М.: Наука, 1982.
5. **Кроновер Р. М.** Фракталы и хаос в динамических системах. Основы теории. М.: Постмаркет, 2000.
6. **Павлов А. Н., Анищенко В. С.** Мультифрактальный анализ сложных сигналов // Успехи физических наук. 2007. Т. 177. № 8. С. 859—876.
7. **Hentschneell H., Procaccia I. P.** The infinite number of generalized dimensions of fractals and strange attractors // Physics D. 1983. V. 8, № 3. P. 435—444.
8. **Grassberger P.** Generalized dimensions of strange attractors // Phys. Lett. A. 1983. V. 97. P. 227—230.
9. **Grassberger P., Procaccia I., Hentschell H.** On the characterization of chaotic motions // Lect. Notes Phys. 1983. V. 179. P. 212—221.
10. **Динамические системы-2.** Современные проблемы математики. Фундаментальные направления. М.: ВИНТИ, 1985.
11. **Майоров В. В., Тимофеев Е. А.** Статистическая оценка обобщенных размерностей // Математические заметки. 2002. Т. 71, вып. 5. С. 697—712.
12. **Семенов В. Ю.** Новый метод локализации корней трансцендентных уравнений // Компьютерная математика. 2004. № 1. С. 143—151.
13. **Ортега Дж. Рейтболдт В.** Итерационные методы решений нелинейных систем уравнений со многими неизвестными. М.: Мир, 1975. 568 с.
14. **Денис Дж., Шнибель Р.** Численные методы безусловной оптимизации и решения нелинейных уравнений. М.: Мир, 1988.
15. **Семенов В. Ю.** Метод нахождения всех действительных некротных корней системы нелинейных уравнений // Журнал вычисл. математики и мат. физики. 2007. Т. 47, № 9. С. 1486—1496.
16. **Бахвалов Н. С.** Численные методы (анализ, алгебра, обыкновенные дифференциальные уравнения). М.: Наука, 1975.
17. **Takens F.** Dynamical systems and turbulence. Berlin: Springer, 1981.
18. **Шустер Г.** Детерминированный хаос. Введение. М.: Мир, 1988.
19. **Moon F. C.** Chaotic and fractal dynamics. New York: John Wiley & Sons Inc, 1992.
20. **Riedi R. H.** An improved multifractal formalism and self-similar measures // J. Math. Anal. Appl. 1995 V. 189. P. 462—490.
21. **Дремин И. М., Иванов О. В., Нечитайло В. А.** Вейвлеты и их использование // Успехи физических наук. 2001. Т. 171, № 5. С. 465—501.
22. **Muzy J. F., Bacry E., Arneodo A.** // Phys. Rev. Lett. 1991. V. 67. P. 3515.
23. **Muzy J. F., Bacry B., Arneodo A.** // Int. J. Bifurcat Chaos. 1994. V. 4. P. 245.
24. **Halsey C. et al.** // Phys Rev. A. 1986. V. 33. P. 1141.

УДК 004.89

Я. Е. Львович, д-р техн. наук, проф.,
зав. каф. САПРИС,

Е. Н. Королев, канд. техн. наук, доц.,
e-mail: korolev73@mail.ru

Воронежский государственный
технический университет

Преимущества объектно-ориентированного подхода для проектирования систем обучения в интеллектуальных системах принятия решений

Описываются основные преимущества объектно-ориентированного подхода при проектировании сложных систем обучения. Предлагаются некоторые графические конструкции языка визуального моделирования систем обучения, применяемые для построения концептуальных, логических и графических моделей сложных систем обучения в интеллектуальных системах принятия решений, отражающих как статические, так и динамические аспекты функционирования систем обучения.

Ключевые слова: проектирование систем обучения, объектно-ориентированные технологии в обучении, визуальное моделирование систем обучения, Enterprise-компоненты обучения

Традиционное построение систем обучения представляется в виде алгоритма, под которым понимается некоторое предписание выполнить точно определенную последовательность действий, направленных на достижение заданной цели, представленной как результат обучения. Такой способ построения обучающих систем сталкивается с трудностями по мере повышения сложности и размерности таких систем, особенно для систем с изменяющимися функциональными требованиями, систем инвариантных к применяемым методам обучения, систем, ориентированных на обучение сложных процессов в интеллектуальных системах принятия решений.

Решить этот комплекс проблем можно, используя технологию объектно-ориентированного проектирования (ООП) применительно к проектированию систем обучения. При этом основны-

ми единицами построения системы обучения на этапе концептуализации, проектирования и программирования будут являться объекты, между которыми можно строить различные отношения, что позволит использовать основные преимущества объектно-ориентированных систем. Каждый объект будет описан классом. Среди классов можно выделить классы для представления моделей данных и классы, реализующие методы обучения. Модели данных задают основные требования к знаниям по каждому конкретному разделу. Применение принципа наследования позволит строить определенные иерархии классов от более общих к более частным, что может задавать общее направление последовательности обучения от более простого к более сложному. Кроме того, беря в качестве базового некоторый готовый класс, можно на его основе строить новые цепи обучения, наследуя от него другие классы. Принцип инкапсуляции позволит скрыть внутреннее устройство класса. Таким образом, класс, содержащий некоторые методы обучения, в зависимости от реализуемого им интерфейса делает видимыми те или иные методы обучения, которые могут понадобиться на разных этапах обучения. Свойство полиморфизма позволит объектам принимать различные внешние формы, т. е. действия, выполняемые определенными методами обучения, могут отличаться в зависимости от того, к кому применяется данный метод. Эти свойства позволяют строить функционально-инвариантные модели обучения.

Благодаря применению принципов ООП система обучения не будет представляться как последовательность predetermined действий, а станет событийно-управляемой. В этом случае процесс обучения будет представлять собой некий цикл ожидания определенных событий, которые должны быть инициированы обучаемым. При наступлении определенного события, например ответа на вопрос, система обучения выходит из состояния ожидания и реагирует соответствующим образом, что связано с генерацией следующих событий. Применение такой методологии требует на этапе анализа и проектирования системы обучения определить основные классы, их свойства и методы.

На этапе анализа предметной области и формулировки требований к системе обучения определяются функции, которые должна выполнять система обучения. Применение методологии объектно-ориентированного анализа и проектирова-

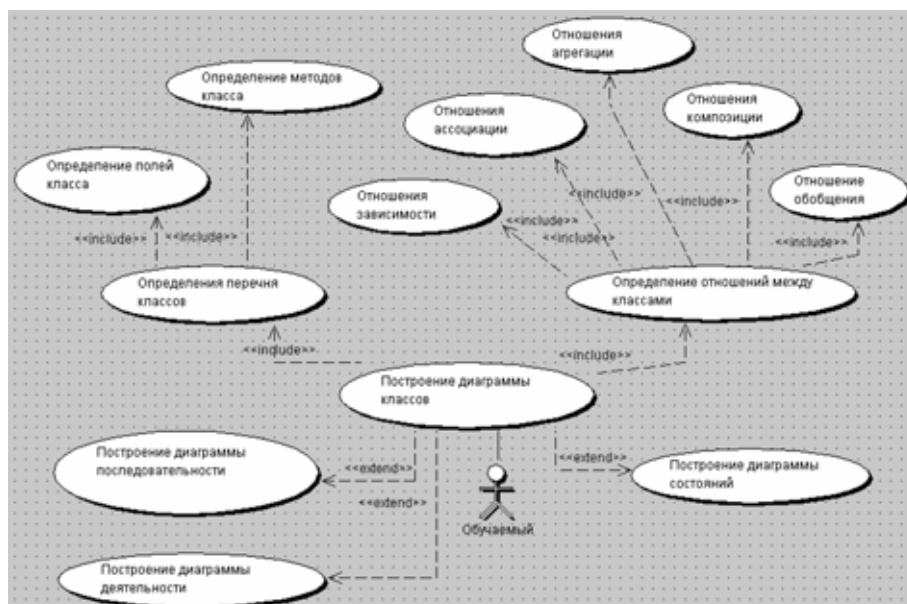
ния для построения систем обучения требует разработки различных средств концептуализации и визуализации моделирования таких систем. Процесс разработки адекватных моделей для сложных систем обучения САПР и их последующего конструктивного использования требует не только применения методологии системного анализа, но и наличие соответствующих средств для фиксации результатов моделирования системы обучения и их документирования. Для построения моделей должны быть разработаны серьезные теоретические методы, основанные на развитии математических и логических средств моделирования, а также предложены различные формальные и графические нотации.

В качестве основы для языка визуального моделирования сложных систем обучения предлагается язык UML, который является простым и мощным средством моделирования и может быть эффективно использован для построения концептуальных, логических и графических моделей сложных систем самого различного целевого назначения.

В рамках языка UML все представления о модели сложной системы обучения будем фиксировать в виде разного рода диаграмм, которые отражают функциональные, статические и динамические аспекты работы системы обучения. Процесс объектно-ориентированного проектирования системы обучения неразрывно связан с процессом построения этих диаграмм. Каждая диаграмма служит для генерации физических компонентов системы обучения. К таким диаграммам относятся диаграммы вариантов использования, диаграммы классов, диаграммы последовательности.

Диаграмма вариантов использования системы обучения САПР служит для определения основных проектных операций, проектных процедур и отношений между ними. Эта диаграмма является исходной для генерации модели обучения, структура которой позволит ей быть функционально-инвариантной для применяемых к ней методов обучения.

Представленная на рисунке диаграмма вариантов использования описывает такую проектную процедуру как "построение диаграммы классов" и определяет для нее основные проектные операции, такие как "определение перечня классов" и "определение отношений между классами". Каждая проектная операция может включать в себя подоперации, которые обозначаются на схеме как



Пример диаграммы вариантов использования системы обучения методом проектирования программного обеспечения

"include", либо указывает возможное расширение данной операции, которое обозначается на схеме как "extend". Определив на диаграмме основные задачи, которые должна решать система для обучения пользователя решению, такой проектной процедуры как "построение диаграммы классов", необходимо создать ту часть системы обучения, которая будет реализовать данную модель, причем она должна быть инвариантна применяемым к ней методам обучения. В терминах понятий объектно-ориентированного анализа и проектирования данная модель представляется в виде класса с набором свойств и методов, имеющего отношения с другими классами. К свойствам данного класса будут относиться обозначенные проектные операции. К методам данного класса будут относиться методы доступа к этим свойствам. Данный класс должен реализовать отношение наследования, т. е. он должен наследоваться от класса, представляющего собой модель обучения проектной процедуре, которая предшествует выполнению проектной процедуры по построению диаграммы классов. Пример части сгенерированной модели для представленной выше диаграммы вариантов использования будет выглядеть следующим образом:

```
public class ClassDiagram implements ClassModel
extends UseCaseDiagram{
    ClassModel listClass;
    ClassModel attributeClass;
    ClassModel methodClass;
    ClassModel relationship;
    ClassModel associationRelationship;
    ClassModel extendRelationship;
```

```

ClassModel generalizationRelationship;
ClassModel includeRelationship;
public Classmodel getListClass()
{ return listClass; }
public void setListClass(Classmodel listClass)
{ this.listClass = listClass; }

```

...

Класс *ClassDiagram* представлен как класс, наследуемый от класса *UseCaseDiagram*, что определяет класс *UseCaseDiagram* базовым по отношению к классу *ClassDiagram*. Следовательно, обучение методам построения "диаграммы классов" должно выполняться после того, как будут изучены методы построения "диаграммы вариантов использования", так как создание экземпляра класса *ClassDiagram* невозможно без создания экземпляра класса *UseCaseDiagram*, что автоматически выстраивает цепочку обучения в соответствующей последовательности. Каждый из заданных классов реализует интерфейс *ClassModel*, который задает некоторый стандарт представления методов, что позволит данным моделям быть функционально инвариантными к методам, их обрабатывающим.

Сформированные функционально-инвариантные модели определяют те проектные операции и процедуры, которые необходимо выполнить для получения определенного проектного решения. Данная модель содержит информацию функционально-инвариантную по отношению к тем методам, которые будут к ним применяться. То есть архитектура и правила построения модели обеспечивают возможность применения к ним универсальных методов, что позволит модели оставаться постоянной, независимо от того, какие методы обучения решению какой-либо проектной процедуры будут использоваться. А различные варианты методов обучения могут быть предложены в зависимости от особенностей обучаемого или промежуточных результатов обучения.

Для обеспечения функциональной инвариантности данной модели задаются некоторые правила объявления методов обучения выполнению проектных операций и самих операций, которые представляются как поля функционально-инвариантной модели. Тогда разработанная модель будет включать в себя знание о методах, к ней применяемых, и не будет зависеть от их реализации.

После построения диаграмм "вариантов использования" и генерации соответствующих им функционально-инвариантных моделей обучения необходимо определить те классы, которые будут реализовывать методы обучения обозначенным проектным процедурам.

Для определения методов обучения формируется диаграмма классов. Каждый класс представляется как *Enterprise*-компонент обучения, реализован-

ный в соответствии с принципами построения EJB-компонент технологии построения распределенных информационных систем J2EE. Эти компоненты выполняются под управлением соответствующего контейнера *Enterprise*-компонент, который должен реализовать в себе некоторый набор сервисов, таких как поддержка жизненного цикла моделей обучения, поддержка транзакций в обучении, поддержка математического аппарата по многоальтернативному выбору направления и последовательности обучения в рамках заданных моделей обучения и построенных между ними отношений.

Для задания вариантов последовательности обучения в рамках одной модели обучения необходимо построение диаграмм последовательности, которые будут служить основой для генерации аспектных моделей, задающих некую последовательность выполнения заданных методов.

Таким образом, использование предложенных объектно-ориентированных технологий для проектирования систем обучения позволит отказаться от традиционного метода построения систем обучения, предписывающих выполнение последовательности predetermined действий. Система обучения, построенная на объектно-ориентированных технологиях, станет событийно-управляемой, легко проектируемой и расширяемой, функционально-инвариантной интеллектуальной системой с возможностью расширения функциональных требований как в процессе проектирования, так и в процессе обучения. Система обучения будет состоять из *Enterprise*-компонентов обучения, инкапсулирующих в себе методы обучения и методы жизненного цикла компонентов и моделей обучения, содержащих информацию об объекте обучения. Сам объект хранит в себе информацию о тематически близких к нему объектах, что позволяет отказаться от традиционного построения процесса обучения в виде алгоритма, и формировать отдельные *Enterprise*-компоненты обучения, которые будут автоматически встраиваться в систему. Использование принципов объектно-ориентированного подхода позволит эти компоненты легко модифицировать без изменения остальных компонентов, а также легко применять их для повторного использования.

Список литературы

1. Рамбо Дж., Блаха М. UML 2.0. Объектно-ориентированное моделирование и разработка. 2-е изд. СПб: Питер, 2007. 544 с.
2. Амриш К. И., Ахмед Х. З. Разработка корпоративных Java-приложений с использованием J2EE и UML. 2002; Вильямс.
3. Aldawud O., Elrad T., Bader A. A UML Profile for Aspect-Oriented Modeling // In proc. of Aspect-Oriented Programming Workshop at OOPSLA. 2001.
4. Stein D., Hanenberg S., Unland R. Designing Aspect-Oriented Crosscutting in UML // In proc. of Workshop on Aspect-Oriented Modeling with UML at AOSD. 2002.

Е. Р. Пантелеев, д-р техн. наук, проф.,
Я. Э. Карпов, студент,
 Ивановский государственный
 энергетический университет,
 e-mail: erp@poks.ispu.ru

Разработка и интерпретация решения расчетных задач в среде Web-обучения

Рассматриваются вопросы создания комплекса методов и средств разработки и интерпретации решения расчетных задач в среде Web-обучения. Обоснованы модель представления задачи в виде раскрашенной сети Петри, а также выбор стандартов сетевой реализации комплекса, обеспечивающий возможность его повторного использования в любой операционной и программной среде. Приведен пример описания и интерпретации процесса решения расчетной задачи в предметной области электромеханики.

Ключевые слова: Web-обучение, расчетная инженерная задача, средства разработки и интерпретации

Компетентность специалиста определяется наличием знаний и навыков, необходимых для эффективного решения задач в заданной предметной области. Согласно [1], задача существует тогда, когда требуется перейти от одного состояния объектов предметной области к другому, если имеется более одного варианта перехода и не все возможные варианты очевидны. Приведенное определение требует дальнейшей детализации для того, чтобы его можно было конструктивно использовать при создании компьютерных систем обучения. Во-первых, любая инженерная задача может быть представлена совокупностью более мелких единиц деятельности (это следует из наличия нескольких вариантов перехода из исходного состояния в целевое). Назовем такую единицу деятельности *инженерной операцией* и определим ее как элементарное действие по переводу объектов задачи из одного состояния в другое. Операция может заключаться в применении расчетного соотношения к переменным задачи, выполнении элементарных построений в инженерной графике, операций сборки машиностроительного изделия или оперативных переключений в энергоустановках.

Во-вторых, на множестве операций задачи определено отношение частичного порядка. Это означает, что возможность применения операций обусловлена текущим состоянием задачи. Например, расчет высоты подъема тела можно провести

только после того, как будет известна вертикальная составляющая скорости. Нельзя подавать напряжение на якорную обмотку двигателя с последовательным возбуждением, если отсутствует момент нагрузки на его валу. В общем случае возможно параллельное выполнение нескольких операций: для расчета высоты подъема тела после расчета вертикальной составляющей можно воспользоваться либо уравнением равноускоренного движения, либо законом сохранения энергии.

В-третьих, возможны ресурсные ограничения на выполнение операции, например, ограничение времени выполнения или ограничение на значения переменных. Таким образом, под *инженерной задачей* будем понимать задачу перевода объектов предметной области из заданного начального состояния в заданное конечное состояние путем применения к ним допустимых операций в допустимой последовательности и с учетом определенных для этих операций ограничений.

Разработка компьютерных обучающих систем для формирования навыков решения инженерных задач предполагает формализацию описания:

- релевантных задаче объектов и операций предметной области;
- самой задачи, включая:
 - ♦ содержательную составляющую (определение начального и конечного состояний задачи и допустимых вариантов преобразования первого состояния во второе; определение методов интерпретации и оценки действий пользователя и содержания учебных воздействий в контексте возможных состояний задачи и действий пользователя);
 - ♦ презентационную составляющую (определение формы визуализации состояний задачи и инструментов для изменения этого состояния пользователем; определение способов регистрации действий пользователя).

Сложность создания подобных систем определяет актуальность разработки повторно используемых компонентов описания инженерных задач. В то же время известные реализации подобных систем (Andes [2] — в области элементарной физики, SQL Tutor [3] — в области языка запросов SQL) используют модели, которые не выделяют инвариантные компоненты описания задачи, что и определяет актуальность исследований в этой области.

В статье рассматриваются методы конструирования описания и интерпретации решения расчетных задач как частного случая задач инженерных, которые обеспечивают возможность повторного использования предложенных методов при создании компьютерных обучающих систем в любой программной и операционной средах.

Учитывая наличие двух компонентов описания расчетной задачи (объекты — переменные, операции — расчетные соотношения), а также возможность псевдопараллельного выполнения соотношений, переводящих переменные из одного дискретного состояния в другое, целесообразно использовать в качестве модели задачи раскрашенную сеть Петри (Colored Petri Net, CPN [4]):

$$CPN = \{P, T, H, G, \mu\}.$$

Здесь P — множество позиций (переменных задачи); T — множество переходов (расчетных соотношений задачи); $H: P \times T \rightarrow R_H$ — входная функция смежности позиций и переходов, которая определяет принадлежность переменных расчетным соотношениям; $G: T \times P \rightarrow R_G$ — выходная

функция смежности переходов и позиций, которая определяет переменные — результаты выполнения операции; $\mu: S \rightarrow P$ — функция разметки, отображающая множество значений S переменных задачи на множество позиций CPN.

Возможность применения расчетного соотношения в данной интерпретации совпадает с условием активности перехода (наличие значений переменных во всех входных позициях), а результат вычисления этого соотношения (новое значение переменной) приводит к изменению разметки CPN.

В качестве примера применения формализма CPN рассмотрим сетевое представление задачи расчета параметров и составляющих напряжения короткого замыкания трансформатора ТМ-5600/35 [5, пример 17—11] (рис. 1). Данная сеть наглядно показывает связь между переменными и расчетными соотношениями задачи, а также допустимые последовательности выполнения расчета. В частности, расчет значений $U_{1кф}$ и $P_{кф}$ может быть выполнен в любой последовательности, а расчет r_k — только после расчета $P_{кф}$.

Повторное использование компонентов конструирования и интерпретации расчетных задач предполагает их встраивание в существующие и вновь разрабатываемые системы компьютерного обучения, а это становится возможным лишь при условии стандартизации программных интерфейсов и форматов представления расчетных соотношений.

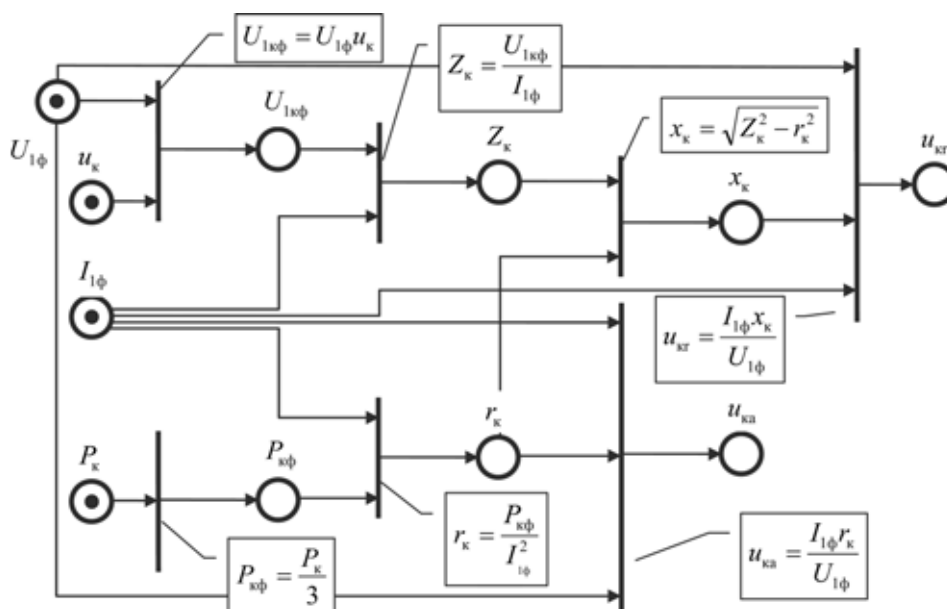


Рис. 1. Сетевая модель задачи расчета параметров короткого замыкания трансформатора:

$U_{1ф}$ — фазное напряжение на стороне высокого напряжения; $U_{1кф}$ — фазное напряжение короткого замыкания на стороне высокого напряжения; $I_{1ф}$ — ток фазы; u_k — коэффициент падения напряжения; P_k — мощность короткого замыкания; $P_{кф}$ — мощность короткого замыкания на фазу; Z_k — полное сопротивление короткого замыкания фазы; r_k — активная составляющая сопротивления короткого замыкания фазы; x_k — реактивная составляющая сопротивления короткого замыкания фазы; $u_{ка}$ — активная составляющая напряжения короткого замыкания; $u_{кз}$ — реактивная составляющая напряжения короткого замыкания

Реализованная в рамках предлагаемого подхода идея стандартизации программного интерфейса Web-приложений основана на использовании языка описания Web-сервисов WSDL [6]. Для того чтобы клиент мог воспользоваться Web-сервисом (по сути — глобально доступной библиотекой подпрограмм), ему достаточно иметь в своем распоряжении описание этого сервиса на языке WSDL и среду программирования, которая поддерживает вызовы Web-сервисов (PHP, Java, ASP. NET, Delphi).

Наиболее приемлемым средством стандартизации записи расчетных соотношений в Web-приложениях является язык разметки математических выражений — MathML*. Он предоставляет как средства разметки представления (рис. 2, а), необходимые для визуального конструирования математических выражений, так и средства разметки содержания (рис. 2, б) для интерпретации и вычисления этих выражений.

Комплекс средств конструирования описаний и интерпретации процесса решения расчетных задач, основанный на описанном выше модельном представлении и стандартах, включает в себя:

- Web-интерфейс авторских средств редактирования задач и вопросов на содержательном уровне (исходные и расчетные переменные,

* <http://www.raleigh.ru/MathML/MathML2>.

```

<mrow>
  <msup>
    <mfenced>
      <mrow>
        <mi>a</mi>
        <mo>+</mo>
        <mi>b</mi>
      </mrow>
    </mfenced>
    <mn>2</mn>
  </msup>
</mrow>

```

а) разметка представления

```

<mrow>
  <apply>
    <power>
      <plus>
        <ci>a</ci>
        <ci>b</ci>
      </plus>
    </power>
    <cn>2</cn>
  </apply>
</mrow>

```

б) разметка содержания

Рис. 2. Пример разметки представления и содержания выражения $(a + b)^2$ в MathML

расчетные соотношения) и на уровне визуального представления;

- Web-сервис доступа к базе расчетных задач для отображения вопросов задачи на компьютере студента, а также обработки ответов студента в контексте состояния данных задачи.

Авторские средства редактирования визуального представления задачи позволяют определить, заполнить и выполнить предварительный просмотр макета клиентской страницы в том виде, в каком ее увидит студент (с точностью до значений случайно генерируемых значений исходных данных для расчета).

Следующий этап — редактирование вопросов (рис. 3, см. третью сторону обложки) — предполагает определение входных и выходных переменных, а также расчетного соотношения для каждого вопроса. В список входных переменных расчетного соотношения могут входить как исходные данные задачи, так и расчетные переменные. Для расчетных переменных окно редактирования содержит лишь имя переменной и наименование единицы измерения (значение не указывается). Для исходных данных задачи кроме имени и единицы измерения переменной необходимо также указать ее тип, значение или диапазон значений.

Рассмотрим создание переменных для расчетного соотношения к вопросу "Рассчитать мощность потерь короткого замыкания на фазу". Это переменные "Мощность короткого замыкания" (P_K) и "Количество фаз" (m). Для исходной переменной значение может быть задано либо константой, либо случайной величиной из заданного диапазона. Для этого под значение переменной отведено два поля. Если заполнены оба, то перед началом решения задачи значение переменной генерируется из заданного промежутка (рис. 4, см. третью сторону обложки), иначе значение считается константой. Редактирование переменных комплексного типа выполняется аналогично, отдельно для действительной и мнимой части.

Редактирование вопроса завершается вводом расчетного соотношения в окно визуального редактора WebEQ (рис. 5, см. третью сторону об-

ложки). В результате повторения описанных операций редактирования для каждого вопроса задачи формируется ее сетевая модель (см. рис. 1).

Интерфейс Web-сервиса интерпретации процесса решения задачи представлен WSDL-файлом*, описывающим предоставляемые Web-сервисом функции "Получить задачу" (getTask) и "Оценить ответ" (sendAnswer). Web-сервис интерпретации решения задачи может быть подключен из любой среды разработки, поддерживающей технологию Web-сервисов (PHP, ASP.NET, Java, Delphi). Наивысший уровень автоматизации процедуры подключения сервиса предоставляет среда ASP.NET. Разработчику приложения-клиента Web-сервиса в этой среде для подключения к сервису даже не требуется прибегать к программированию. Достаточно указать адрес (URL) WSDL-файла и получить список доступных функций.

При первом обращении сервис выполняет персональную параметризацию исходных данных задачи, т. е. генерацию случайных данных из диапазона, заданного при разработке задачи. Так, в задаче расчета параметров короткого замыкания трансформатора методом случайной генерации будут инициализированы следующие параметры: $U_{1ф}$ — напряжение на фазе; $I_{1ф}$ — сила тока; u_k — коэффициент падения напряжения; P_K — мощность.

Цикл решения задачи представляет собой последовательность пар вопросов типа "что делать?" и "как делать?" на текущем шаге решения. Вопрос "что делать?" предъявляется студенту в виде задания на выбор одного из оставшихся шагов решения задачи, соответствующих активным переходам CPN. Для правильно выбранного шага должны быть известны значения всех исходных данных для расчета. Вопрос "как делать?" представляет собой задание, предусматривающее ответ в виде расчетного соотношения, соответствующего выбранному шагу задания (рис. 6, см. третью сторону обложки). Сервис отправляет клиенту страницу представления вопроса в соответствии с его типом и разработанным шаблоном.

Сценарий заканчивает работу, как только список доступных вопросов становится пустым. В качестве оценки решения Web-сервис возвращает отношение числа правильных ответов к общему числу вопросов задачи.

Таким образом, в результате выполненных исследований разработана модель представления расчетной задачи в виде раскрашенной сети Петри, метод ее интерпретации, Web-интерфейс авторских средств разработки задачи и Web-сервис ее интерпретации. Применение разработанных методов и средств позволяет авторам учебно-ме-

* Этот файл хранится в реестре сервисов ИГЭУ и доступен для использования любому клиентскому приложению.

тодических материалов и программистам-разработчикам компьютерных средств обучения использовать предложенные решения многократно, независимо от программной и операционной платформы разработки. Средства разработки и интерпретации расчетных задач интегрированы в систему Web-обучения ГИПЕРТЕСТ 2.0 [7] и испытаны на примерах задач по курсам основ электротехники и электрических машин и аппаратов.

Список литературы

1. Крик Э. Введение в инженерное дело. М.: Энергия, 1970. 176 с.
2. Gertner A. and VanLehn K. Andes: A Coached Problem Solving Environment for Physics / Gauthier G., Frasson C., and Van-

Lehn K. (Eds.) // Proceedings of 5th International Conference on Intelligent Tutoring Systems, Springer-Verlag. 2000. P. 133–142.

3. Mitrovic A. Learning SQL with a Computerized Tutor // Proc. 29th SIGCSE Tech. Symposium 1998. P. 307–311.

4. Jensen K. Coloured Petri Nets: A High-level Language for System Design and Analysis. / G. Rozenberg (ed.) // Advances in Petri Nets 1990, Lecture Notes in Computer Science. Springer-Verlag. 1991. Vol. 483. P. 342–416.

5. Костенко М. П., Пиотровский Л. М. Электрические машины: Учебник для вузов в 2 ч. Изд. 3-е. Ч. 1. Машины постоянного тока. Трансформаторы. Л.: Энергия, 1973.

6. Ньюкомер Э. Веб-сервисы: XML, WSDL, SOAP и UDDI Understanding Web Services: XML, WSDL, SOAP and UDDI. Сер.: Для профессионалов. СПб.: Питер, 2003, 256 с.

7. Пантелеев Е. Р., Ковшова И. А., Малков И. В., Пекунов В. В., Первовский М. А., Юдельсон М. В. Среда разработки программ дистанционного обучения и профильного тестирования ГИПЕРТЕСТ: инструментальные средства // Информационные технологии. 2001. № 8. С. 34–40.

УДК 007;004.3

В. В. Сафронов, д-р техн. наук, проф.,
ОАО "КБ Электроприбор", г. Саратов,
e-mail: svv@kbep.ru

Метод ветвей и границ в прямой и обратной задачах оптимального распределения времени на изучение цикла дисциплин

Рассмотрены прямая и обратная постановки задач обоснованного выбора времени на изучение дисциплин при подготовке специалистов с высшим профессиональным образованием. Для решения таких задач предлагается использовать метод ветвей и границ. Раскрыты особенности построения оценок на уровнях ветвления. Приведены численные примеры.

Ключевые слова: государственное образовательное учреждение, метод ветвей и границ, оценки, время на изучение цикла дисциплин

Введение

Важнейшими документами для Государственных образовательных учреждений высшего профессионального образования являются Государственные образовательные стандарты (ГОС). При формировании ГОС одной из актуальных является задача рационального (оптимального) распределения времени на изучение цикла дисциплин (например, общие гуманитарные и социально-экономические дисциплины, общие математические и естественно-научные дисциплины, специ-

альные дисциплины и т. п.). В свою очередь, актуальна и задача рационального (оптимального) распределения времени на изучение дисциплин внутри каждого цикла. Не менее важной представляется задача оптимального распределения времени на изучение отдельных тем данной дисциплины. Не теряя общности, изложение будем вести для задачи оптимального распределения времени по дисциплинам внутри цикла.

1. Математическая постановка прямой и обратной задач

Допустим, на основе экспертного опроса получены сведения о том, сколько часов занятий необходимо для изучения отдельных дисциплин на оценки "удовлетворительно", "хорошо", "отлично". Пусть известно суммарное время, выделяемое на все рассматриваемые дисциплины цикла. Необходимо так распределить время, отводимое на отдельные дисциплины, чтобы сумма баллов, получаемых студентами, была максимальной при условии, что мы укладываемся в отведенный бюджет времени и по каждой дисциплине оценка не ниже, чем "удовлетворительно".

Рассмотрим математическую постановку задачи. С этой целью введем обозначения:

- $d_j(x_j)$ — время, необходимое для изучения j -й дисциплины в целях получения оценки "удовлетворительно" ($x_j = 1$), "хорошо" ($x_j = 2$), "отлично" ($x_j = 3$), $j = \overline{1, n}$, где n — число дисциплин анализируемого цикла;
- x_j — номера возможных значений времени $d_j(x_j)$, $x_j \in \{1, 2, 3\}$;

- $C_j(x_j)$ — значения оценок (в баллах) по j -й дисциплине, в частности,

$$C_j(x_j = 1) = 3; C_j(x_j = 2) = 4;$$

$$C_j(x_j = 3) = 5 \quad \forall j = \overline{1, n};$$

- n — число дисциплин анализируемого цикла;
- D^0 — время, выделяемое на изучение дисциплин данного цикла (константа ограничений).

С учетом введенных обозначений сформулируем нашу задачу.

Найти вектор $X = \{x_j, j = \overline{1, n}\}$, доставляющий максимальное значение функции

$$C = \max_{x_j} \sum_{j=1}^n C_j(x_j), \quad (1)$$

при выполнении ограничения

$$\sum_{j=1}^n d_j(x_j) \leq D^0. \quad (2)$$

Не менее важной является *обратная*, по отношению к (1), (2), задача. Допустим, на основе экспертного опроса получены сведения о том, сколько часов занятий необходимо для изучения отдельных дисциплин на оценки "удовлетворительно", "хорошо", "отлично". Пусть определен суммарный (средний) балл изучения рассматриваемого цикла дисциплин.

Необходимо так распределить время, отводимое на изучение отдельных дисциплин, чтобы время на изучение всего цикла дисциплин, было бы минимальным при условии, что суммарный (средний) балл — не меньше заданного значения и по каждой дисциплине оценка не ниже, чем "удовлетворительно".

Математическая постановка задачи заключается в следующем. Найти вектор $X = \{x_j, j = \overline{1, n}\}$, доставляющий минимальное значение функции

$$D = \min_{x_j} \sum_{j=1}^n d_j(x_j) \quad (3)$$

при выполнении ограничения

$$\sum_{j=1}^n C_j(x_j) \geq C^0, \quad (4)$$

где C^0 — заданный суммарный балл изучения цикла дисциплин.

2. Метод решения задачи

Задачи (1), (2) и (3), (4) относятся к области целочисленного программирования. Для их решения используем метод ветвей и границ (МВГ) [1–4].

Не снижая общности, решение рассмотрим для задачи (1), (2). Общая идея метода заключа-

ется в замене полного перебора частичным перебором на основе построения верхних оценок и исключения бесперспективных вариантов.

С этой целью для каждого элемента x_j ($j = \overline{1, n}$) вычисляют верхние оценки, исходя из следующего рассуждения: какое, в лучшем случае, будет значение целевой функции, если в решение войдет рассматриваемый элемент.

Далее из всех элементов в решение включают такой элемент r_j , который имеет наибольшую оценку $T_j(r_j)$, и из него проводят дальнейшее ветвление. Вновь вычисляют оценки уже для элементов $(j + 1)$ -го уровня ветвления и т. д.

На последнем уровне ветвления имеем решение (пока возможно приближенное), т. е. набор элементов, вошедших в систему, обеспечивающих заданное значение времени, отводимого на изучение всех дисциплин, и значение целевой функции $T^0(r_n)$. Далее проверяем решение на оптимальность и, если возможно, уточняем его.

Проверка на оптимальность заключается в следующем.

Поднимаемся на $(n - 1)$ -й уровень ветвления. Смотрим, есть ли вершины (концевые, неветвленные) с оценками, лучшими, чем значение $T^0(r_n)$. Если есть, то проводим уточнение путем ветвления из вершины с наилучшей оценкой.

В противном случае поднимаемся на уровень $(n - 2)$ и т. д. Если все уровни будут просмотрены, то решение заканчиваем.

В соответствии с МВГ первоначально рассчитываем верхнюю границу целевой функции T^* , которая показывает, что мы можем получить в самом лучшем случае от системы с точки зрения значения целевой функции. Если на n -м уровне ветвления значения $T^0(r_n)$ и T^* совпадают, то решение заканчиваем.

Таким образом, из качественного описания МВГ следует, что здесь существенным являются три момента [1]:

1) вычисление оценки границы решения для каждого элемента вершины;

2) определение правила ветвлений;

3) отсеечение бесперспективных вершин.

Для вычислений верхней границы решается некоторая оптимизационная задача более простой структуры, чем исходная. При определении верхней границы применяют два основных подхода:

- замену дискретной задачи непрерывной (на шаге $k = v$, когда происходит переход через ограничение), оптимальное решение которой берется в качестве верхней границы;
- неучет части ограничений.

Правила вычисления верхней границы во многом зависят от специфики задачи и искусства исследователя. Чем точнее верхняя граница, чем она ближе к оптимальному решению, тем больше от-

бракуется неперспективных вершин и меньше времени уходит на поиск оптимального решения. Вместе с тем, за более точную оценку приходится "платить" усложнением алгоритма и временем, необходимым для ее нахождения.

При необходимости задача оптимизации в соответствии с МВГ может быть решена с относительной погрешностью ε , которая определяется из выражения [1]

$$\varepsilon = \frac{T^* - T^0(r_n)}{T^*}.$$

Если $\varepsilon \leq \varepsilon_0$, где ε_0 — заданное значение погрешности, то решение заканчиваем, в противном случае решение уточняем.

Таким образом, важнейшие особенности метода ветвей и границ — это определение оценок границы решения, правила ветвления и отсечение бесперспективных вариантов.

Рассмотрим, с учетом сделанных замечаний, использование идей МВГ для решения задачи (1), (2).

Обозначим $T_\beta(x_\beta)$ — верхняя граница (оценка) вершины $x_\beta(x_\beta = \overline{1, 3})$ β -го уровня ветвления. После вычисления оценок всех вершин x_β в решение включаем такой элемент r_β , для которого справедливо

$$T_\beta(r_\beta) = \max_{x_\beta} T_\beta(x_\beta), \quad x_\beta \in \{\overline{1, 3}\}.$$

При достижении уровня $\beta = n$ получаем оценку решения T^0 ($T^0 = T_n(r_n)$). Если при этом $T^0 = T^*$, то полученное решение оптимально. В противном случае переходим к проверке вершин, не вошедших в решение, для которых справедливо

$$T_\beta(x_\beta) > T^0(x = \overline{1, 3}, x_\beta \neq r_\beta, \beta = \overline{1, n-1}).$$

Допустим, что на уровнях ветвления 1, 2, ..., $\beta - 1$ реализации уже включены в решение $R_{\beta-1} = \{r_i, i = \overline{1, \beta-1}\}$. Тогда на β -м уровне ветвления оценку $T_\beta(x_\beta)$ можно определить зависимостью

$$T_\beta(x_\beta) = C_\beta(x_\beta) + \sum_{i=1}^{\beta-1} C_i(r_i) + Q_\beta^0(x_\beta), \quad (5)$$

где $Q_\beta^0(x_\beta)$ — верхняя граница максимального значения суммы

$$Q_\beta(x_\beta) = \max_{x_i} \sum_{i=\beta+1}^n C_i(x_i), \quad (6)$$

при условии

$$\sum_{i=\beta+1}^n d_i(x_i) \leq b(x_\beta), \quad (7)$$

где

$$b(x_\beta) = D^0 - \sum_{i=1}^{\beta-1} d_i(r_i) + d_\beta(x_\beta). \quad (8)$$

В выражении (5) первое слагаемое учитывает балл рассматриваемого элемента, второе — баллы элементов, включенных в решение на предыдущих уровнях, третье — оценку оставшейся части из числа не рассмотренных элементов. Очевидно, что $Q_\beta^0(x_\beta) \geq Q_\beta(x_\beta)$.

Таким образом, как следует из (5), задача отыскания верхней границы сводится к определению $Q_\beta^0(x_\beta)$.

Для определения верхней границы целевой функции T^* полагаем

$$R_\beta = \emptyset, \quad b(x_\beta) = D^0, \quad \sum_{i=1}^{\beta} C_i(x_i) = 0. \quad (9)$$

Таблица 1

Исходные данные

№ дисциплины	$C_j (x_j = 1)$	$C_j (x_j = 2)$	$C_j (x_j = 3)$	$d_j (x_j = 1), \text{ ч}$	$d_j (x_j = 2), \text{ ч}$	$d_j (x_j = 3), \text{ ч}$
1	3	4	5	60	70	80
2	3	4	5	100	120	130
3	3	4	5	80	90	100
4	3	4	5	60	70	90
5	3	4	5	70	90	110
6	3	4	5	40	50	70
7	3	4	5	90	110	130
8	3	4	5	60	90	110
9	3	4	5	50	70	80
10	3	4	5	30	60	70
11	3	4	5	72	90	100
12	3	4	5	200	220	260
13	3	4	5	300	340	360
14	3	4	5	160	170	190
15	3	4	5	40	60	90
Общее время, ч				1412	1960	

Таблица 2

Результаты решения задачи

№ дисциплины	Оптимальное время (ч) на изучение дисциплины и оценка (балл) при $D^0 = 1600$ ч	Оптимальное время (ч) на изучение дисциплины и оценка (балл) при $D^0 = 1800$ ч
1	80 (5)	80 (5)
2	130 (5)	130 (5)
3	100 (5)	100 (5)
4	90 (5)	90 (5)
5	70 (3)	110 (5)
6	50 (4)	70 (5)
7	90 (3)	130 (5)
8	60 (3)	60 (3)
9	80 (5)	80 (5)
10	30 (3)	70 (5)
11	100 (5)	100 (5)
12	200 (3)	220 (4)
13	300 (3)	300 (3)
14	170 (4)	190 (5)
15	40 (3)	60 (4)
Сумма баллов	59	69
Средний балл	3,93	4,6
Общее время, ч	1590	1790

Таблица 3

Обобщенные результаты решения обратной задачи

Суммарный балл, C^0	Средний балл	Минимальное время, необходимое для изучения цикла дисциплин, ч
60	4	1610
63	4,2	1670
66	4,4	1730
67,5	4,5	1770
69	4,6	1790

Таблица 4

Результаты решения обратной задачи

№ дисциплины	Оптимальное время (ч) на изучение дисциплины при среднем балле, равном 4,2	Оптимальное время (ч) на изучение дисциплины при среднем балле, равном 4,4
1	80 (5)	80 (5)
2	130 (5)	130 (5)
3	100 (5)	100 (5)
4	70 (4)	70 (4)
5	70 (3)	70 (3)
6	50 (4)	50 (4)
7	90 (3)	130 (5)
8	60 (3)	60 (3)
9	80 (5)	80 (5)
10	70 (5)	70 (5)
11	100 (5)	100 (5)
12	200 (3)	220 (4)
13	300 (3)	300 (3)
14	190 (5)	190 (5)
15	90 (5)	90 (5)
Сумма баллов	63	66
Минимальное суммарное время, ч	1670	1730

Тогда $T^* = Q_\beta^0(x_\beta)$ — верхняя граница решения задачи (6)—(8) с учетом (9).

Для определения верхней границы используем градиентный метод с оценкой, предложенной П. Колешаром [3].

Условием окончания выбора является выполнение на шаге $k = v$ соотношения $b^{v-1} < b(x_\beta) \leq b^v$, т. е. переход через ограничение. В качестве оценки, казалось, можно бы взять C^{v-1} . Однако в этом случае часть ограничения, а именно $(b(x_\beta) - D^{v-1})$, будет недоиспользована. Поэтому, рассуждая: "что будет, если ограничение используем полностью, отказываясь временно на этом шаге от условия целочисленности", получим оценку $f^0(x_\beta) = C^v + h^v G$, где $h^v = \frac{C^v - 1 - C^v}{D^v - D^{v-1}}$, $G = (D^v - b(x_\beta))$. Заметим, что если $b(x_\beta) = D^{v-1}$, то $f^0(x_\beta) = C^{v-1}$.

Метод решения задачи (3), (4) аналогичен изложенному выше, только на уровнях ветвления отыскивают не верхнюю, а нижнюю границу со всеми вытекающими вычислительными особенностями.

3. Численный пример

Исходные данные, сформированные экспертной группой, были использованы для решения задачи распределения учебного времени по пятнадцати дисциплинам цикла (табл. 1).

Из табл. 1 видно, что минимальный бюджет времени, который может быть выделен на изучение рассматриваемых дисциплин, составляет 1412 ч, а максимальный, позволяющий получить по всем дисциплинам оценку "отлично", — 1960 ч. Задача решалась для значений констант ограничений из данного интервала времени. Результаты решения задачи для двух значений ограничений приведены в табл. 2.

Полученные результаты позволяют обоснованно определить время, необходимое для изучения дисциплин внутри рассматриваемого цикла.

Рассмотрим решение *обратной* задачи. Для различных значений C^0 (среднего балла изучения цикла дисциплин) с использованием метода ветвей и границ получены минимальные значения времени на изучение цикла дисциплин (табл. 3).

Результаты решения задачи для двух значений среднего балла приведены в табл. 4. В скобках указан получаемый балл.

Анализ результатов решения позволяет сделать обоснованный выбор времени, необходимого для изучения дисциплин внутри рассматриваемого

цикла. Заметим, что время решения каждой из задач составляет единицы секунд.

Заключение

Рассмотрена важная в прикладном плане задача оптимального распределения времени по дисциплинам. Задача решена на основе применения метода ветвей и границ. Использование предлагаемого подхода позволяет:

- обоснованно распределить время на дисциплины цикла данной специальности, решая прямую либо обратную задачу;
- оценить, как влияет бюджет времени на успеваемость студентов;
- корректно осуществлять реструктуризацию рабочих программы;

- решать и иные задачи обоснования времени, выделяемого на изучение цикла дисциплин ГОС, отдельных тем конкретной дисциплины с учетом требований, предъявляемых к специалисту с высшим профессиональным образованием.

Список литературы

1. **Алексеев О. Г.** Комплексное применение методов дискретной оптимизации. М.: Наука. Гл. ред. физ.-мат. лит., 1987. 248 с.
2. **Корбут А. А., Финкельштейн Ю. Ю.** Дискретное программирование / Под ред. Д. Б. Юдина. М.: Наука. Гл. ред. физ.-мат. лит., 1969. 386 с.
3. **Kolesar P. J.** A branch-and-bound algorithm for knapsack problem // Manag. Sci. 1967. V. 13. № 9. P. 723—735.
4. **Сафронов В. В.** Основы системного анализа: методы многовекторной оптимизации и многовекторного ранжирования. Саратов: Научная книга, 2009. 329 с.

ДИСКУССИОННЫЙ КЛУБ

УДК 004.03

И. М. Гуревич, канд. техн. наук, ст. науч. сотр.,
Институт проблем информатики РАН,
г. Москва, e-mail: iggurevich@gmail.ru

Информация как универсальная неоднородность

До сих пор нет общепринятого определения информации, хотя известно много десятков определений. Это обусловлено, прежде всего, большим числом и разнообразием предметных областей, в которых используется понятие "информация". Понятие "информация" используется практически во всех областях науки, культуры, экономики. В настоящей статье предпринимается очередная попытка дать общее определение информации. Точнее говоря, автор формализует и обосновывает универсальность одного из известных определений. Приводятся примеры из различных предметных областей.

Ключевые слова: однородность, неоднородность, информация, информационная энтропия, информационная дивергенция, объем информации

Примеры известных определений информации

Приведем примеры определений, изложенные в [1, 2]. "Информация есть знания, переданные кем-то другим или приобретенные путем собственного исследования или изучения", "информация — это сведения, содержащиеся в данном сообщении и рассматриваемые как объект передачи, хранения и обработки", "информация в обы-

денном смысле — это сведения, известия, в научно-технических приложениях — то, что имеет на себе сигнал", "информация означает порядок, коммуникация есть создание порядка из беспорядка или по крайней мере увеличение степени той упорядоченности, которая существовала до получения сообщения", "информация — это обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств", "информация... — одно из свойств предметов, явлений, процессов объективной действительности, созданных человеком управляющих машин, заключающееся в способности воспринимать внутреннее состояние и воздействие окружающей среды и сохранять определенное время результаты его; передавать сведения о внутреннем состоянии и накопленные данные другим предметам, явлениям и процессам", "информация — объективное содержание связи между взаимодействующими материальными объектами, проявляющееся в изменении состояний этих объектов", "информация есть текущие данные о переменных величинах в некоей области деятельности, систематизированные сведения относительно основных причинных связей, которые содержатся в знании как понятия более общего класса, по отношению к которому информация является подчиненной", "информация есть знание о каком-то особом событии, случае или о чем-либо подобном", "информацией являются все те данные о внешнем мире, которые мы получаем как путем непосредственного воз-

действия на наши органы чувств окружающих предметов и явлений, так и опосредованным путем через книги, газеты, рассказы других людей", "информацией называется всякое сообщение или передача сведений о чем-либо, что заранее не было известно", "информация в самом общем случае — это разнообразие, которое один объект содержит о другом, это взаимное, относительное разнообразие. С позиций теории отражения информация может быть представлена как отраженное разнообразие, как разнообразие, которое отражающий объект содержит об отраженном", "информация есть отражение в сознании людей объективных причинно-следственных связей в окружающем нас реальном мире", "информация — это содержание процессов отражения", "информация не тождественна отражению, а есть лишь его инвариантная часть, поддающаяся определению, объективированию, передаче", "информация — это философская категория, рассматриваемая наряду с такими, как пространство, время, материя. В самом общем виде информацию можно представить как сообщение, т. е. форму связи между источником, передающим сообщение, и приемником, его принимающим", "информация (от лат. *informatio* — ознакомление, разъяснение, представление, понятие) 1) сообщение, осведомление о положении дел, сведения о чем-либо, передаваемые людьми; 2) уменьшаемая, снимаемая неопределенность в результате получения сообщений; 3) сообщение, неразрывно связанное с управлением, сигналы в единстве синтаксических, семантических и прагматических характеристик; 4) передача, отражение разнообразия в любых объектах и процессах (неживой и живой природы)", "информация (от латинского *informatio*) — осведомление", "информация... есть план строения клетки и, следовательно, всего организма", "информация есть некий алгоритм", "совокупность приемов, правил или сведений, необходимых для построения оператора, будем называть информацией. Под словом "оператор" здесь понимается некое стороннее воздействие на систему, изменяющее спонтанный ход событий". "Большое число похожих и непохожих друг на друга определений понятия "информация" означает, что общепринятого определения еще нет. Более того, нет даже четкого понимания сути этого явления, хотя потребность в нем уже назрела. Сейчас область применимости информационного подхода существенно расширилась. Понятие "информация" используется при исследовании практически всех процессов самоорганизации (в частности биологической эволюции)". Таковы типичные и широко используемые определения информации.

В последнее время многие ученые используют определения информации, введенные Л. Бриллю-

эном и Д. С. Чернавским, определения опирающиеся на физические понятия. Л. Бриллюэн: "Мы введем теперь различие между двумя видами информации: 1) свободная информация (*free information*), возникающая, когда возможные случаи рассматриваются как абстрактные и не имеющие определенного физического значения; 2) связанная информация (*bound information*), возникающая, когда возможные случаи могут быть представлены как микросостояния физической системы".

Д. С. Чернавский: "Информация есть случайный и запомненный выбор одного варианта из нескольких возможных и равноправных. Здесь слово "случайный" выделено, поскольку оно относится к процессу (способу) выбора и потому сужает область применимости определения. Вообще говоря, выбор может быть и не случайным (подсказанным), в этом случае говорят о рецепции информации. Случайный выбор соответствует генерации (то есть спонтанному возникновению) информации. Вообще говоря, выбор может и не запоминаться (то есть тут же забываться). Такой выбор будем называть микроинформацией. Запомненный выбор (в отличие от незапоминаемого) будем называть макроинформацией. Во всех информационных процессах используется макроинформация (запоминаемая)".

Н. С. Кардашев отметил, что наряду с материей и энергией Вселенная содержит, включает в себя и информацию [3]. Информация является неотъемлемой частью Вселенной. По определению автора [4, стр. 35], информация — это устойчивая определенное время неоднородность: "Под информацией мы будем понимать устойчивые определенные время неоднородности произвольной физической природы. Тем самым буква в книге, атом, молекула, элементарная частица, звезда, чертеж, рисунок, вспаханное поле, лес и другие неоднородности содержат и несут информацию. Вместе с тем флуктуации плотности, напряжения не являются информационно-содержащими".

По аналогичному мнению ряда ученых (В. М. Глушкова [5], Г. В. Встовского "Мы предлагаем и обсуждаем еще один возможный ответ: информация определяется нарушенной симметрией в рассматриваемых системах, все известные меры являются мерами нарушенных симметрий. Все объекты, системы, явления или процессы, изучаемые той или иной частной дисциплиной, неизбежно имеют некоторые ограничения на их симметрию (начальные или граничные условия, геометрическая структура, и т. д.), поэтому они должны содержать в себе некоторую нарушенную симметрию, т. е. информацию" [6], К. К. Колина [7]), по своей физической сущности информация — это неоднородности в распределении материи и энергии.

Основополагающий принцип квантовой механики А. Цайленгера [8] постулирует, что элементарная физическая система несет один бит информации — этот постулат определяет информацию как одну из характеристик физических систем (наряду с массой, зарядом, энергией, ...).

Информация — это устойчивая определенное время неоднородность

Определение, данное автором, использовалось главным образом для физических систем. Покажем, что данное определение имеет общий характер и может использоваться для систем произвольного вида.

Приведенное определение информации основывается на следующих понятиях:

- неоднородность;
- устойчивая неоднородность;
- устойчивая определенное время неоднородность.

"Кто же не знает, что такое однородность и неоднородность? Эти понятия используются буквально во всех областях человеческой деятельности: однородные и неоднородные функции в математике, однородная атмосфера в физике, однородная деформация в технике, однородные и неоднородные смеси в химии и технологии. Так что же это за свойство, называемое однородностью и что такое неоднородность? — Попробуем сформулировать: однородность — это..., впрочем, давайте заглянем в Большую Советскую Энциклопедию. Ищем: "обзорность", "огнеупорность", "одаренность", "одушевленность"...всякие свойства есть, а "однородности" нет. Посмотрим на букву "н": "необратимость", "неодолимость", "неодушевленность". Все есть кроме "неоднородности". Нет этих терминов и в других энциклопедиях — в Философской, Технической, Географической и др., включая Большую Британскую Энциклопедию. Термины "неоднородность", "неоднородное тело" и т. п. отсутствуют в Толковом словаре живого великорусского языка Владимира Даля; нет их и в словаре Ожегова, нет в Политехническом словаре, не нашлось для них места и в десятичном издании "Научно-техническая терминология", а также в созданных Академией Наук СССР: в Словнике толкового словаря естественно-научных терминов и в Логическом словаре-справочнике. Бесполезно также рыться в Толковом словаре математических терминов, в Философском словаре и многих, многих других" [9]. Такова судьба понятий "однородность" и "неоднородность" — всем они очевидны и все их используют, не задумываясь об определениях.

Дадим определения однородности и неоднородности. Рассмотрим множество M , состоящее из эле-

ментов m . Если элементы m одинаковы, тождественны (не отличаются друг от друга), то множество M однородно.

Если среди элементов m есть неодинаковые (не тождественные) (отличающиеся друг от друга), то множество M неоднородно.

Рассмотрим множество M , состоящее из элементов m , определяемых параметрами α . Если параметры α для всех элементов m одинаковы, тождественны (не отличаются друг от друга), то множество M однородно. Если среди параметров, α элементов m есть неравные (не тождественные) (отличающиеся друг от друга), то множество M неоднородно. Под множеством, как обычно, понимается набор, совокупность, собрание каких-либо объектов, называемых его элементами, обладающих общими для всех их характеристическими свойствами. В конце XIX века Георг Кантор определил множество как "единое имя для совокупности всех объектов, обладающих данным свойством". Эти объекты называются элементами множества. Множество объектов, обладающих свойством $A(x)$, обозначается $\{x | A(x)\}$. Если некое множество $Y = \{x | A(x)\}$, то $A(x)$ называется характеристическим свойством множества Y [10].

Устойчивость неоднородности — сохранение неоднородности, ее структуры, характеристик.

Время устойчивости неоднородности — время существования (жизни) неоднородности, время сохранения неоднородности, ее структуры, характеристик.

Используя данное автором определение информации и следуя Д. С. Чернавскому, можно дать классификацию различных типов информации. В таблице сведены определения различных типов информации.

С понятием информация связано понятие "носитель информации". Носитель информации — физическая система, объект, содержащий неоднородность.

Приведем примеры определения носителя информации. Носителем информации является все, что может воспринимать, хранить и передавать информацию (человек, животное, жесткий диск, растение, книга и т. д.) [10].

Носитель информации — материальный объект, используемый для закрепления и хранения на нем речевой, звуковой или изобразительной информации, в том числе в преобразованном виде [11].

Неоднородности могут использоваться как для хранения информации (предпочтительно, фермионы), так и для передачи информации (предпочтительно, бозоны) [12].

Примечание 1. *Флуктуации* — случайные отклонения от среднего значения физических величин, характеризующих систему из большого числа частиц; вызываются тепловым движением

Определения различных типов информации

Тип информации	Время существования неоднородности	Тип неоднородности
Классическая информация	Время существования неоднородности бесконечно $t_n = \infty$	Абсолютно устойчивая неоднородность
Макроинформация	Время существования неоднородности не меньше времени существования системы $t_n \geq t_s$	Существенно устойчивая неоднородность
Информация	Время существования неоднородности меньше времени существования системы, но больше времени протекания процессов в системе, времени оценки, измерения характеристик неоднородности $t_{n1} < t_n < t_s$	Устойчивая неоднородность
Микроинформация	Время существования неоднородности меньше времени протекания процессов в системе, времени измерения характеристик неоднородности $t_n < t_{n1}$	Неустойчивая неоднородность
Примечание. t_n — время существования неоднородности; t_s — время существования системы; t_{n1} — время оценки, измерения неоднородности.		

частиц или квантово-механическими эффектами; не являются устойчивыми определенное время и поэтому в стационарных системах не несут макроинформации. В нестационарных системах (например, во Вселенной) флуктуации могут порождать макроинформацию.

Примечание 2. Следует отметить, что предложенное автором определение информации характеризуется не только общностью, но и простотой. Оно использует только одно из самых общих математических понятий — понятие множества.

Некоторые классы однородностей и неоднородностей

Перечислим некоторые классы однородностей и неоднородностей [10].

- Естественные (природные):
 - ♦ неорганические, в том числе физические;
 - ♦ биологические;
 - ♦ экологические;
 - ♦ другие.
- Искусственные:
 - ♦ материальные;
 - ♦ абстрактные (идеальные);
 - ♦ абстрактно-материальные.
- Смешанные:
 - ♦ социально-технологические;
 - ♦ организационно-технические;
 - ♦ социально-экономические;
 - ♦ другие.

Приведем примеры.

Физические однородности: пространство; время; вакуум и т. д.

Иерархия физических неоднородностей: калибровочные поля; фундаментальные частицы; элементарные частицы; атомы; молекулы; плазма; газы; жидкости; твердые тела (кристаллы); звезды; галактики; вселенная.

Иерархия геологических неоднородностей: Земля; ядро; мантия; земная кора; платформы; океаны и моря; материки; горы; равнины; впадины; горные породы; минералы.

Лингвистические однородности: текст из пробелов, одинаковых символов.

Иерархия лингвистических неоднородностей: буквы; слова; предложения; тексты; книги; библиотеки.

Технические однородности как фон, сцена: элементы одного типа.

Иерархии технических неоднородностей:

элемент; связь; подсистема; система; модуль; блок; устройство; изделие; комплекс; сеть; модуль; блок; подпрограмма; программа; комплекс; сеть.

Биологические однородности: хромосомы в клетках одного организма.

Иерархии биологических неоднородностей:

аминокислота; белок; нуклеотиды; ДНК; РНК; клетка; одноклеточный организм; вирус; бактерия; многоклеточный организм; популяция.

Социально-экономические однородности: деньги.

Иерархия социально-экономических неоднородностей: индивид; семья; группа; нация; производственная единица; отрасль; народное хозяйство; страна; цивилизация.

Каждый уровень иерархии неоднородностей произвольного типа содержит различные элементы x_i . На каждом уровне иерархии неоднородностей можно определить вероятность p_i реализации каждого элемента x_i (в дискретном случае) и функцию распределения $F(x)$, плотность распределения $p(x)$ элементов x (в непрерывном случае).

Информационная энтропия по Шеннону — универсальная мера неоднородности

Универсальной мерой физической неоднородности — информации — является информационная энтропия по Шеннону [13, 14].

Энтропия H дискретной случайной величины определяется в битах (в натах):

$$H = - \sum_{i=1}^n p_i \log_2 p_i \quad \left(H = - \sum_{i=1}^n p_i \ln p_i \right).$$

Энтропия H непрерывной случайной величины определяется в битах (натах):

$$H(x) = -\int p(x) \log_2 p(x) dx \quad (H(x) = -\int p(x) \ln p(x) dx).$$

Важно отметить, что энтропия по И. Фон Нейману $S = \rho \ln \rho$, где ρ — матрица плотности [15], не может использоваться в качестве меры неоднородности, поскольку она равна нулю для имеющего структуру чистого состояния, хотя физики, как правило, для описания и исследований квантовых систем используют именно энтропию Неймана.

Информационная дивергенция — универсальная мера наличия неоднородности

Само наличие неоднородности удобно оценивать информационной дивергенцией [16, 17].

Пусть случайная величина x описывается функцией (плотностью) распределения $P(x)$. Отличие случайной величины с функцией (плотностью) распределения $P(x)$ от случайной величины с функцией (плотностью) распределения $Q(x)$ оценивается информационной дивергенцией $D(P/Q)$ распределения P относительно распределения Q :

$$\begin{aligned} D(P/Q) &= -\int P(x) \log_2 \frac{P(x)}{Q(x)} dx = \\ &= -\int P(x) \log_2 P(x) dx + \int P(x) \log_2 Q(x) dx. \end{aligned}$$

Наличие неоднородности, задаваемой распределением P , будем оценивать информационной дивергенцией $D(P/R)$ распределения P относительно равномерного распределения R :

$$\begin{aligned} D(P/R) &= -\int P(x) \log_2 \frac{P(x)}{R(x)} dx = \\ &= -\int P(x) \log_2 P(x) dx + \int P(x) \log_2 R(x) dx, \end{aligned}$$

где $P(x)$ — распределение, характеризующее искомую неоднородность; $R(x)$ — равномерное распределение на интервале $0 \leq x \leq a$:

$$R(x) = \begin{cases} 0 & \text{при } -\infty < x \leq 0; \\ \frac{1}{a} & \text{при } 0 < x \leq a; \\ 0 & \text{при } a < x \leq \infty. \end{cases}$$

Если $R(x)$ — равномерное распределение на интервале $0 \leq x \leq a$ и $P(x)$ также определена на интервале $0 \leq x \leq a$, то информационная дивергенция

$$D = -\int_0^a P(x) \log_2 \frac{P(x)}{\frac{1}{a}} dx = -\int_0^a P(x) \log_2 (aP(x)) dx.$$

Аналогично оценивается неоднородность информационной дивергенцией в многомерном случае.

Информационная дивергенция имеет важные свойства.

Утверждение 1. Если распределение $P(x)$ равномерно, т. е. $P(x) \equiv R(x)$, то информационная дивергенция $D(P(x)/R(x))$ в координатах x равна нулю: $D(P(x)/R(x)) = 0$.

При переходе от координат $x = (x_1, \dots, x_n)$ к координатам $y = (y_1, \dots, y_n)$

$$\begin{aligned} y_1 &= y_1(x_1, \dots, x_n); \\ &\dots\dots\dots \\ y_i &= y_i(x_1, \dots, x_n); \\ &\dots\dots\dots \\ y_n &= y_n(x_1, \dots, x_n); \end{aligned}$$

плотности распределения связаны следующим образом [13]:

$$P(y) = P(x) J \frac{(x_1, \dots, x_n)}{(y_1, \dots, y_n)},$$

где $J \frac{(x_1, \dots, x_n)}{(y_1, \dots, y_n)} = \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} = \left| \frac{\partial x_i}{\partial y_j} \right|$ — якобиан преобразования, обратного к преобразованию $y = y(x)$.

При переходе от координат $x = (x_1, \dots, x_n)$ к координатам $y = (y_1, \dots, y_n)$ получаем $P(y) = P(x) J \frac{(x_1, \dots, x_n)}{(y_1, \dots, y_n)}$ и $R(y) = R(x) J \frac{(x_1, \dots, x_n)}{(y_1, \dots, y_n)}$. Следовательно, дивергенция в новых координатах определяется выражением

$$\begin{aligned} D(P(y)/R(y)) &= -\int P(y) \log_2 \frac{P(y)}{R(y)} dy = \\ &= -\int P(x) J \left(\frac{x}{y} \right) \log_2 \frac{P(x) J \left(\frac{x}{y} \right)}{R(x) J \left(\frac{x}{y} \right)} dx = \\ &= -\int P(x) J \left(\frac{x}{y} \right) \log_2 \frac{P(x)}{R(x)} dx = \\ &= -\int P(x) J \left(\frac{x}{y} \right) \log_2 P(x) dx + \int P(x) J \left(\frac{x}{y} \right) \log_2 R(x) dx. \end{aligned}$$

Утверждение 2. Если распределение $P(x)$ равномерно: $P(x) \equiv R(x)$, и y есть функция x : $y = y(x)$ ($y_i = y_i(x_1, \dots, x_n)$), то информационная дивергенция $D(P(y)/R(y))$ в координатах y равна нулю:

$$D(P(y)/R(y)) = 0.$$

Утверждение 3. Если y есть функция x : $y = y(x)$ ($y_i = y_i(x_1, \dots, x_n)$), то информационная дивергенция $D(P(x)/R(x))$ распределения $P(x)$ относительно равномерного распределения $R(x)$ и информационная дивергенция $D(P(y)/R(y))$ распределения

$P(y)$ относительно равномерного распределения $R(y)$ равны (не равны) нулю одновременно:

$$D(P(x)/R(x)) = 0 \Leftrightarrow D(P(y)/R(y)) = 0$$

или $D(P(x)/R(x)) \neq 0 \Leftrightarrow D(P(y)/R(y)) \neq 0$.

Несколько интересных примеров неоднородностей

Слова, сообщения. Если использовать английский алфавит, то аааааа — однородное слово, несущее информацию как целое; аааааа, abcdeg — неоднородные слова, содержащие информацию в целом и в отдельных буквах.

Объем информации в слове abcdeg (считаем все буквы равновероятными) равен $6 \cdot 4,007 = 28,2$ бит.

Замечание. В множестве однородных слов различной длины неоднородность порождается такой характеристикой слова, как его длина.

Бит. Бит — это система, имеющая два состояния 0, 1. Пусть p_1, p_2 — вероятности нахождения бита в состояниях 0, 1.

Информационная энтропия бита

$$N = -(p_1 \log_2 p_1 + p_2 \log_2 p_2).$$

Максимальная информационная энтропия бита достигается при $p_1 = p_2 = 1/2$ и равна 1 бит.

Кубит. Квантовый бит — кубит (q -бит) [18–19]. Это система, имеющая два выделенных (базисных) состояния ($|0\rangle, |1\rangle$) и состояния ψ , представимые в виде линейных комбинаций выделенных состояний

$$\psi = c_1|0\rangle + c_2|1\rangle,$$

где c_1, c_2 — произвольные комплексные коэффициенты, сумма квадратов модулей которых равна единице: $|c_1|^2 + |c_2|^2 = 1$.

Неопределенность (информационная энтропия) q -бита с двумя выделенными состояниями равна

$$N = -(|c_1|^2 \log_2 |c_1|^2 + |c_2|^2 \log_2 |c_2|^2),$$

где $|c_1|^2, |c_2|^2$ — вероятности нахождения q -бита в состояниях $|0\rangle, |1\rangle$. Максимальная неопределенность (информационная энтропия) q -бита с двумя выделенными состояниями достигается при $|c_1|^2 = |c_2|^2 = 1/2$ и равна 1 бит.

Квантовый компьютер. Это система, состоящая из n q -битов, имеющая 2^n выделенных (базисных) состояний x_i и состояния ψ , представимые в виде линейных комбинаций выделенных состояний [18–19]:

$$\psi = \sum c_i x_i.$$

Выделенные состояния образованы прямым произведением выделенных состояний n q -битов;

c_i — произвольные комплексные коэффициенты, сумма квадратов модулей которых равна 1:

$$\sum |c_i|^2 = 1.$$

Состояние квантового компьютера представляется в виде вектора-строки $|\psi\rangle$ (кет-вектора).

Размерность пространства состояний квантового компьютера равна 2^n . Неопределенность q -бита с двумя выделенными состояниями равна

$$N = - \sum_{i=1}^n |c_i|^2 \log_2 |c_i|^2.$$

Максимальная неопределенность квантового компьютера, состоящего из n q -битов с двумя выделенными состояниями каждый, достигается при $|c_i|^2 = 2^{-n}$ и равна n бит.

Неопределенность и информация наблюдаемых* и состояний квантовых систем

Приведем определение неопределенности, введенное автором в 1989 г. в работе [4]. Используем описание физической системы волновой функцией. Согласно квантовой механике, для системы, находящейся в чистом состоянии, часть параметров a может задаваться точно, значения другой части параметров b при этом задаются функцией или плотностью распределения $P(b) = |\psi_a(b)|^2$, где $\psi_a(b)$ — волновая функция, описывающая физическую систему в b -представлении (волновая функция, описывающая наблюдаемую b). Система, находящаяся в чистом состоянии $\psi_a(b)$, имеет внешнюю неопределенность наблюдаемой b , равную информационной энтропии К. Шеннона для непрерывных случайных величин [13, 14] $N_a(b) = - \int |\psi_a(b)|^2 \log_2 |\psi_a(b)|^2 db$ бит. Как правило, внешняя неопределенность определяется на волновых функциях, зависящих от пространственно-временных координат.

Внутренняя неопределенность определяется на волновых функциях (амплитудах вероятности) в гильбертовых пространствах, описывающих состояние системы. Рассмотрим объект в состоянии $|\alpha\rangle$, описываемом амплитудами вероятностей $\langle \beta_i | \alpha \rangle$ $|\alpha\rangle = \sum |\beta_i\rangle \langle \beta_i | \alpha \rangle$. Величина $\langle \beta_i | \alpha \rangle^2$ есть вероятность обнаружения наблюдаемой β , объекта, находящегося в состоянии $|\alpha\rangle$, в базисном состоянии $|\beta_i\rangle$. Неопределенность наблюдаемой β , объекта, находящегося в состоянии $|\alpha\rangle$ равна информационной энтропии Шеннона для дискретных случайных величин $N = - \sum |\langle \beta_i | \alpha \rangle|^2 \log_2 |\langle \beta_i | \alpha \rangle|^2$ бит. Неопределенность определяется состоянием, в

* Наблюдаемой в квантовой механике называют любую физическую величину, которую можно измерить, причем результатами эксперимента обязательно должны являться действительные числа.

котором находится система и наблюдаемой, описывающей систему. В общем случае неопределенность разных наблюдаемых системы, находящейся в заданном состоянии, может быть различна. Если наблюдаемая фиксирована, то можно говорить о неопределенности состояния квантовой системы. Точно также можно говорить о неопределенности состояния квантовой системы в случае задания системы координат из априорных соображений.

Полная неопределенность системы равна сумме внешней и внутренней неопределенности.

Примечание 3. Информационная энтропия служит одновременно для оценки неопределенности и информации. "Эта величина измеряет также количество неопределенности, содержащейся в этом эксперименте, т. е. количество неопределенности до бросания кости относительно того, каков будет его результат. Наконец, эта величина измеряет информацию, содержащуюся в этом эксперименте, или количество информации, получаемой в результате бросания. Тот факт, что случайность и неопределенность имеют естественную общую меру, неудивителен. Вследствие "формулы" прирост информации = устраненная неопределенность представляется разумным, что неопределенность и информация должны измеряться с помощью одной и той же функции" [20].

В отличие от энтропии, используемой в статистической физике и характеризующей неопределенность состояний, в которых может находиться физическая система, информационная энтропия характеризует неопределенность конкретного состояния, точнее говоря, неопределенность наблюдаемой системы, находящейся в некотором состоянии. Во избежание неоднозначности терминологии для обозначения свойства "неопределенность" и его количественной меры — информационной энтропии — будем использовать один термин — "неопределенность". Введенное определение неопределенности придает широко используемому в квантовой механике понятию неопределенности точный *информационный смысл*.

Смешанное состояние $\psi_s(x)$ определяется как смесь чистых состояний $\psi_i(x)$. Обозначим p_i вероятность реализации чистого состояния $\psi_i(x)$. Неопределенность наблюдаемой в смешанном состоянии системы N_s равна сумме средней неопределенности наблюдаемой в чистых состояниях плюс энтропия системы (неопределенность состояний)

$$N_s = N + S.$$

Здесь N — средняя неопределенность наблюдаемой в (чистых) состояниях системы $N = \sum_i p_i N_i$; N_i — неопределенность наблюдаемой в i -м (чистом) со-

стоянии системы; S — энтропия системы (неопределенность состояний системы) $S = -\sum_i p_i \ln p_i$.

Основные свойства неопределенности (информации). Приведем основные свойства.

- Неопределенность физических систем (наблюдаемых) может принимать любые значения от $-\infty$ до $+\infty$, $-\infty \leq N \leq +\infty$.
- Классические объекты, приборы, обладающие точными значениями всех своих параметров, имеют неопределенность, равную $-\infty$ (не ноль!). Точнее говоря, неопределенность, равную $-\infty$, имеют состояния, соответствующие точным значениям некоторых наблюдаемых в представлении этих же наблюдаемых.
- Свободные частицы имеют пространственную неопределенность, равную $+\infty$.
- Неопределенность физической системы S , состоящей из невзаимодействующих подсистем $S_1, S_2, \dots$, равна сумме неопределенностей этих подсистем (аддитивность): $N_s = N_{s1} + N_{s2} + \dots$
- Неопределенность физической системы S , состоящей из взаимодействующих подсистем S_1, S_2 , не превосходит суммы неопределенностей этих подсистем: $N_{s1} + N_{s2} + \dots \geq N_s$. Можно сказать, что в системе, состоящей из взаимодействующих подсистем, образуется дефект неопределенности: $\Delta N_s = N_{s1} + N_{s2} + \dots - N_s$.
- Если у физической системы S , находящейся в смешанном состоянии, число состояний M , в которых может находиться система, велико так, что $\ln M \gg \max N_i$, то неопределенность системы можно считать приблизительно равной энтропии системы.

Начальные неоднородности (информация) Вселенной. Приведенные выше свойства информационной дивергенции дают возможность оценить начальные неоднородности Вселенной [21–23].

Утверждение 4. В начальные моменты времени неоднородности обычной материи во Вселенной существовали.

Доказательство. Доказательство существования начальной неоднородности проведем, используя метод "от противного". Предположим, что в начальный момент времени t_0 во Вселенной не было неоднородностей обычной материи: $P(x, t_0) \equiv R(x)$. Тогда информационная дивергенция в произвольный момент времени $t \geq t_0$ также равна нулю. Это означает, что в произвольный момент времени $t \geq t_0$ во Вселенной также нет неоднородностей обычной материи. Но в настоящее время в нашей Вселенной, очевидно, есть неоднородности обычной материи (скопления галактик, галактики, звезды, планеты, молекулы, атомы, частицы [21]). Следовательно, используя информационный подход, мы методом "от противного" доказали, что в начальные моменты

времени во Вселенной были неоднородности обычной материи.

Примечание 4. Данное утверждение справедливо при любой физической природе неоднородностей, при любом физическом механизме образования неоднородностей, любой физической модели формирования неоднородностей.

Примечание 5. Данный результат дает строгое обоснование утверждению Я. Б. Зельдовича с соавторами. "Сразу же подчеркнем, что какие-то начальные неоднородности Вселенной необходимы, так как на фоне абсолютно однородной Вселенной образование крупномасштабной структуры (галактик, их скоплений) невозможно" [21, стр. 12–13].

Утверждение 5. В начальные моменты времени неоднородности темной материи во Вселенной существовали.

Доказательство аналогично предыдущему.

В темной энергии (вакууме) неоднородности отсутствуют в начальный и последующие моменты времени.

Утверждение 6. В начальный и последующие моменты времени неоднородностей в темной энергии не было. Темная энергия (вакуум) была распределена равномерно

$$R(x) = \begin{cases} 0 & \text{при } -\infty < x \leq 0; \\ \frac{1}{a} & \text{при } 0 < x \leq a; \\ 0 & \text{при } a < x \leq \infty. \end{cases}$$

Информационная дивергенция $D(P(x)/R(x))$ темной энергии в настоящий момент времени равна нулю. Тогда согласно утверждению 3 в начальный и последующие моменты времени информационная дивергенция $D(P(y)/R(y))$ темной энергии равна нулю.

Звезды. Звезды — космологические объекты, отличающиеся друг от друга координатами, физическими характеристиками (массой, размером, температурой, моментом импульса, структурой, составом, ...).

Объем информации, содержащийся в звезде, определяется ее структурой, составом, информацией в составных частях и взаимосвязях составных частей [24–27].

Биологическая информация. Используем сведения, изложенные в [28–29].

Ген — материальный носитель наследственной информации, которую родители передают потомкам во время размножения. Гены — это участки ДНК, несущие какую-либо целостную информацию — о строении одной молекулы белка или одной молекулы РНК.

Генетический код — это свойственный всем живым организмам способ кодирования аминокислотной последовательности белков с помощью последовательности нуклеотидов.

кислотной последовательности белков с помощью последовательности нуклеотидов.

В ДНК используется четыре нуклеотида — аденин (А), гуанин (Г), цитозин (С), тимин (Т), которые в русскоязычной литературе обозначаются буквами А, Г, Ц и Т. Эти нуклеотиды составляют буквы алфавита генетического кода. В РНК используются те же нуклеотиды, за исключением тимина, который заменен похожим нуклеотидом — урацилом, который обозначается буквой У (У в русскоязычной литературе). В молекулах ДНК и РНК нуклеотиды выстраиваются в цепочки и, таким образом, получаются последовательности генетических букв. Объем информации в одном нуклеотиде равен 2 бит.

Для построения белков в природе используется 20 различных аминокислот. Каждый белок представляет собой цепочку или несколько цепочек аминокислот в строго определенной последовательности. Эта последовательность определяет строение белка, а следовательно, все его биологические свойства. Набор аминокислот также универсален для почти всех живых организмов. Объем информации в одной аминокислоте равен 4,32 бит.

Информация передается от нуклеиновых кислот к белку, но не в обратном направлении. Переход генетической информации от ДНК к РНК и от РНК к белку является универсальным для всех без исключения клеточных организмов, лежит в основе биосинтеза макромолекул. Репликации генома соответствует информационный переход ДНК → ДНК. В природе встречаются также переходы РНК → РНК и РНК → ДНК (например, у некоторых вирусов), а также изменение конформации белков, передаваемое от молекулы к молекуле.

Информационная сущность денег. Деньги как информация имеют ряд особенностей [1]. "При наличии денег в количестве M условных единиц можно потратить их на приобретение различных видов товаров. Число различных вариантов трат N равно $N = 2^M$. Количество информации о возможных вариантах покупок $I_N = \log_2 N = M$ ".

Рассмотрим пример. Пусть у нас имеются 10 видов товаров ценою 1 руб. каждый, а у покупателя 3 руб. Тогда число вариантов приобретения товаров равно 10^3 :

$$N = 10^3 = 2^{3,32 \cdot 3}.$$

Заключение

Вышеизложенное позволяет сделать следующие основные выводы.

- Можно утверждать, что информация — это неоднородность произвольной физической природы.

- Все известные автору определения информации в разных предметных областях опираются на различия между рассматриваемыми объектами и фактически (хотя в некоторых случаях неявно) используют понятие неоднородности. Всюду рассматривается несколько различных исходов, альтернатив, результатов выбора.
- Следует отметить, что предложенное автором определение информации характеризуется не только общностью, но и простотой. Оно использует только одно из самых общих математических понятий — понятие множества и понятие неоднородности, характеризующее различие между элементами множества.
- Универсальной мерой неоднородности (информации) является информационная энтропия по Шеннону. Каждая система характеризуется неопределенностью, информацией.
- Само наличие неоднородности оценивается информационной дивергенцией распределения, описывающего неоднородность, относительно равномерного распределения.
- Свойства информационной дивергенции дают возможность строго доказать что в начальные моменты времени неоднородности обычной материи во Вселенной существовали.
- Для информационного описания произвольной предметной области, оценок информационных характеристик рассматриваемых объектов необходимо:
 - ◆ выделить множество рассматриваемых объектов;
 - ◆ описать характеристики объектов;
 - ◆ описать вероятности распределения значений характеристик рассматриваемых объектов;
 - ◆ оценить информационную дивергенцию и информационную энтропию характеристик рассматриваемых объектов.

Список литературы

1. Чернавский Д. С. Синергетика и информация (динамическая теория информации). Изд. 2-е, испр. и доп. М.: Едиториал УРСС. 2004. 288 с.
2. Бриллюэн Л. Наука и теория информации. М.: Гос. изд-во физико-математической литературы. 1960.
3. Кардашев Н. С. Информация, компьютеры и цивилизации во Вселенной. Совместное заседание семинара по космо-

ческой философии НКЦ SETI и секции "Поиски внеземных цивилизаций" НСА РАН. М.: ГАИШ, 7 апреля 2006.

4. Гуревич И. М. Законы информатики — основа исследований и проектирования сложных систем связи и управления: Мет. пособие. М.: ЦООНТИ "Экос", 1989.
5. Глушков В. М. О кибернетике как науке // Кибернетика, мышление, жизнь. М.: Мысль. 1964. С. 53—62.
6. Встовский Г. В. Элементы информационной физики. М.: МГИУ, 2002. С. 9—10.
7. Колин К. К. Природа информации и философские основы информатики // Открытое образование. 2005. № 2 (49). С. 43—51.
8. Zeilinger A. A Foundational Principle for Quantum Mechanics // Foundations of Physic. 1999. N 29 (4). P. 631—643.
9. <http://aleran.ru/?Neodnorodnostmz>.
10. <http://ru.wikipedia.org/wiki/>.
11. [http://www.glossary.ru/cgi-bin/gl_sch2.cgi?RNuxoylrot\(uwsg.oo](http://www.glossary.ru/cgi-bin/gl_sch2.cgi?RNuxoylrot(uwsg.oo)
12. Гуревич И. М. Законы информатики — основа строения и познания сложных систем. Изд. второе, уточненное и доп. М.: Торус Пресс. 2007. 400 с.
13. Шеннон К. Математическая теория связи. Работы по теории информации и кибернетики. М.: Изд-во иностр. лит., 1963. С. 243—332.
14. Стратонович Р. Л. Теория информации. М.: Сов. радио, 1975.
15. Фон Нейман М. Математические основы квантовой механики. М.: Наука, 1964.
16. Чисар И., Кернер Я. Теория информации. М.: Мир, 1985. 400 с.
17. Кульбах С. Теория информации и статистика. М.: Наука. 1967. 408 с.
18. Валиев К. А., Кокин А. А. Квантовые компьютеры: Надежда и реальность. М.-Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика". 2001.
19. Валиев К. А. Квантовые компьютеры и квантовые вычисления // УФН. 2005. Т. 175. № 1.
20. Биллингсгей П. Эргодическая теория и информация. М.: Мир, 1969.
21. Зельдович Я. Б., Новиков И. Д. Строение и эволюция вселенной. М.: Наука, 1975.
22. Долгов А. Д., Зельдович Я. Б., Сажин М. В. Космология ранней Вселенной. М.: Изд. МГУ, 1988. 199 с.
23. Постнов К. А. Лекции по общей астрофизике для физиков. Раздел 12 // Курс кафедры астрофизики и звездной астрономии "Общая астрофизика" (для студентов физического факультета). URL: <http://www.astronet.ru/db/msg/1170612/index.html>.
24. Гуревич И. М. Информационная модель черной дыры // Тр. конф. ВАК-2007. Казань. 2007.
25. Гуревич И. М. Об информационных моделях в космологии. Системы и средства информатики. М.: ИПИ РАН, 2007. Вып. 17. С. 164—183.
26. Gurevich I. ESTIMATION OF INFORMATION VOLUME IN COSMOLOGY OBJECTS // Тр. II Междунар. конф. "Проблемы кибернетики и информатики". Баку, сентябрь 2008.
27. Гуревич И. М. Оценка основных информационных характеристик Вселенной // Информационные технологии. № 12. Приложение. 2008. 32 с.
28. Уотсон Д. Д. Молекулярная биология гена: Пер. с англ. М.: Мир, 1967.
29. Дубинин Н. П. Общая генетика. М.: Наука, 1970.

М. В. Рожкова, независимый специалист
рынков мобильных гаджетов, устройств
и технологий спутниковой
GPS-ГЛОНАСС навигации,
e-mail: rozhkova.m.v@gmail.com

Мои заметки о Mobile World Congress 2010. Тезисно: значимые события и тренды

18.02.2010 закончился очередной **Mobile World Congress** в Барселоне — одна из самых значимых в году выставок телеком-индустрии. По сравнению с прошлым годом в этом году посетителей было меньше (по некоторым оценкам, 40 тысяч против 60 тысяч в 2009 г.). Предлагаю вашему вниманию некую выжимку моих личных впечатлений относительно самых значимых событий выставки в сегменте мобильных гаджетов, событий, которые, на мой взгляд, будут определять вектор развития сегмента в этом году. Подчеркиваю, это "summary" личных впечатлений. Кратко и тезисно. Длинные и подробные отчеты вы в изобилии сможете найти в Интернете на специализированных ресурсах.

1. С большой помпой прошел анонс операционной системы **Windows Mobile 7**. Первые устройства на ней ожидаются лишь к концу 2010 г. Полностью переработана идеология операционной системы, структура меню, доступа к различным функциям. Ничего похожего на "старую" Windows Mobile. Никакой попытки сохранить похожесть с настольной Windows — то, на что Microsoft делал упор ранее.

Новая ОС предъявляет достаточно высокие аппаратные требования (минимальная по функционалу железка — HTC HD2). Поэтому для бюджетных устройств параллельно с WM7 будет дальше развиваться WM 6.5. То есть будут две операционки — для Hi-End и для Middle-Range/Low-End сегментов.

Совместимости Windows Mobile 7 со старым софтом, написанным для Windows Mobile предыдущих версий, практически не будет.

2. На выставке было представлено **много бюджетных и простых сотовых телефонов** (\$20—30). Большое внимание уделяется телефонам для лиц с ограниченными возможностями, также для по-

жилых людей и детей. Налицо новый тренд рынка, новый растущий сегмент подобных устройств.

3. Широтой размаха поразил стенд **Samsung** — один из самых заметных и крупных на выставке.

Одно из главных событий выставки — анонс Samsung Wave на собственной платформе Bada. Дисплей устройства выполнен по технологии Super AMOLED — отличается большими углами обзора и отличной цветопередачей. Картинка реально притягивает полнотой цвета и четкостью. Именно благодаря экрану устройство получилось без преувеличения хитовое — ничего подобного нет ни у одного другого бренда.

Сама операционная система от Samsung, как мне кажется, не несет в себе ничего революционно нового: традиционный упор на пальцеориентированность, "коктейль" из iPhone и Android. Но судя по размаху экспонирования Wave и то, что именно устройство на новой платформе Bada оснащено таким революционным экраном, упор в 2010 г. будет сделан именно на продвижение своей собственной платформы.

Сторонние разработчики ПО говорят, что хотя Bada-SDK и не идеален и не лишен некоторых шероховатостей и недостатков, разрабатывать софт на новую платформу не намного сложнее, чем для Google Android.

4. **Nokia** в выставке не участвовала. Была пресс-конференция компании, посвященная новой платформе (ничего замечательного или экстраординарного в ней нет), но отдельно стоящего стенда Nokia не было. Об утере позиций говорить пока, как мне кажется, рано, но налицо некий провальный тренд. По крайней мере, если сравнивать с активностью вышеупомянутого Samsung.

5. **Мобильный софт**. Самая значимая премьера — SPB Software представил очередную, пятую версию своего хитового приложения — оболочку SPB Mobile Shell для Android, Symbian и Windows Mobile. Еще больший упор на 3D-анимацию, собственная адресная книга с массой интересных возможностей.

6. Подавляющее большинство среди представленных новинок сегмента коммуникаторов-смартфонов — на базе **Android OS**. Очевидно, что на Android бросились все, но также есть основание полагать, что в 2010 г. рынок будет перенасыщен устройствами на Android. Как показывают первые результаты продаж андроидфонов в России, спрос будет существенно отставать от пред-

ложения. Тем более, учитывая полное игнорирование компанией Google (российским офисом, в частности) вопросов продвижения платформы, повышения информированности целевой аудитории о преимуществах платформы — потребители, продавцы на точках довольно часто не понимают, что такое Android, зачем он нужен, в чем его преимущества.

По уровню развития (в частности коммерциализации) рынка стороннего ПО Android пока также довольно сильно отстает от Windows Mobile и iPhone: число доступных приложения в разы меньше, в Android Market ПО невозможно пока приобретать из России, уровень продаж ПО для Android на порядок отстает от уровня продаж ПО для других платформ.

Не исключено, что к концу года будет наблюдаться "обратный откат" производителей к платформе Windows Mobile — определенно намечается некий дефицит моделей с этой ОС. Тем более, если учесть активность Microsoft с появлением WM7.

Самые значимые новинки сегмента коммуникаторов:

Бренд	Windows Mobile	Android	Собственные (проприетарные) ОС
HTC	HTC HD Mini	HTC Desire HTC Legend	HTC Smart
Acer	Acer neoTouch P300 Acer neoTouch P400	Acer beTouch E400 Acer beTouch E110 Acer beTouch Liquid e	—
Samsung	—	—	Samsung Wave
Motorola	—	Motorola Quench Motorola Snapper Motorola Titanium	—
Sony Ericsson	—	Xperia X10 Mini Xperia X10 Mini Pro	—
ZTE	—	ZTE Blade ZTE Racer ZTE Smooth ZTE Mercury	—
Huawei	—	Huawei U8110 Huawei U8220 Huawei U8230	—
Toshiba	Toshiba TG02 Toshiba K01	—	—
Alcatel	—	Alcatel OT-980	—

Как можно заметить, число анонсов устройств на Android в разы больше, чем в Windows Mobile. Общая статистика по числу новинок такова:

- Windows Mobile — 3 компании-вендора, 5 устройств (20% общего числа новинок)
- Android — 7 компаний, 18 устройств (70% общего числа новинок)
- Собственные (проприетарные) ОС — 2 компании, правда самые крупные из производителей коммуникаторов — Samsung и HTC. 2 устройства (менее 10% от общего числа новинок).

Итог: сумеет ли рынок "проглотить" такую порцию Android — вопрос спорный.

Из конкретных устройств мое внимание привлёк новый HTC HD Mini. Эдакая уменьшенная копия HD2 с интересными решениями по дизайну. Похоже, мы присутствуем при рождении новой концепции дизайна коммуникаторов: современный коммуникатор должен быть стильным и снаружи и внутри. HTC попыталась идеально "вылизать" дизайн не только внешнего вида устройства, но и внутренностей — открывая отсек с аккумулятором, можно видеть, что дизайнеры поработали и тут. Это не просто случайный набор электронных компонент, но продуманное и стильное расположение внутренних элементов.

Интересно, что одна из новинок Acer — neoTouch P300 на Windows Mobile — вызвал острый приступ дежа-вю. Дело в том, что абсолютно идентичный аппарат, но только на Android, уже полгода как продается в России под брендом Highscreen Zeus. На данный момент это — самый бюджетный Android-фон в России и самый доступный по цене QWERTY-слайдер среди устройств всех платформ — 9990 рублей. Устройство выпускает компания Foxconn — контрактный сборщик всех Apple iPhone. Видимо, Acer счел аппарат достаточно интересным функционально и достаточно качественным в плане сборки, чтобы поставить на него свой бренд и продавать от собственного имени. Устройство, скорее всего, не появится в России (как мне сказали представители Acer), но если так случится, то у обладателей Acer neoTouch P300 появится возможность "перепрошить" свой аппарат на Android, а у пользователей Highscreen Zeus, скорее всего, возможность "перешить" аппарат на Windows Mobile. Что для пользователя, согласитесь, только плюс.

**Открытое акционерное общество
«Окружной выставочный центр «Югорские контракты»**

628011, Тюменская область, Ханты-Мансийский автономный округ — Югра,
г. Ханты-Мансийск, ул. Студенческая, 19

Тел./факс (3467) 359-596, 359-598

E-mail: ugraexpo@r86.ru, www.yugcont.ru

9—12 июня в городе Ханты-Мансийске в рамках международного IT-форума состоится **VIII межрегиональная конференция «Информационные технологии и решения для «Электронной России» и IX окружная выставка «Информационные технологии XXI века».**

Организаторы: Комитет по информационным ресурсам Ханты-Мансийского автономного округа — Югры.

Цель выставки — обмен опытом в области внедрения и использования современных IT-решений, развитие и укрепление международных и межрегиональных связей, выработка рекомендаций для повышения уровня региональной информатизации.

На выставке «Информационные технологии XXI века» будут представлены новейшие технологии и услуги рынка IT, возможности их применения во всех сферах государственного управления, бизнеса и общественной жизни. Выставка способствует построению в России информационного общества, росту отечественной экономики и повышению жизненного уровня населения.

В рамках деловой программы, включающей пленарные и секционные заседания, «круглые столы» и мастер-классы, будут рассматриваться актуальные вопросы формирования информационного общества, создания «электронного государства», предоставления электронных услуг гражданам, обеспечения информационной безопасности, доступа к официальным информационным ресурсам и др.

Основные разделы проекта:

- Электронные регионы, электронные муниципалитеты
- Подключение государственных органов к компьютерным сетям
- IT — парки
- IT — бизнес-инкубаторы
- Информационно-справочные службы; системы интерактивного взаимодействия населения с органами власти; видеоконференции и телемосты
- Электронный документооборот в органах государственной власти, системы межведомственного документооборота
- Использование информационных технологий для взаимодействия государства с бизнесом; системы электронной отчетности, налогообложения, сбора статистических данных и контроля профиля и статуса предприятий частного бизнеса
- Использование инфокоммуникаций в сфере социального обеспечения и в страховании

- Системы идентификации пользователя, системы защиты информации, электронная цифровая подпись
- Прикладные программы для управления бизнесом
- Телекоммуникации как средство человеческого общения
- Средства развлечения, использование инфокоммуникационных систем доступа к развлекательным программам

Ваше участие в выставке и ее посещение дают Вам возможность за 3 дня выйти на новый уровень развития в Ханты-Мансийском автономном округе — Югре!

5 мая — День для протокольных мероприятий: в первый день выставки к осмотру экспозиции будут приглашены представители Правительства автономного округа и Администрации г. Ханты-Мансийска:

- демонстрация новейших отечественных и зарубежных инноваций и технологий сферы ИТ;
- содействие в продвижении новых технологий на рынок округа;
- содействие в обновлении базы поставщиков региона для госзаказа;
- демонстрация широкого спектра предоставляемых товаров и услуг для государственного заказчика;
- укрепление имиджа участников на рынке Ханты-Мансийского автономного округа в период зарождения новой политической, экономической, социальной и культурной истории округа.

6 мая — День для профессионалов отрасли: день направлен на сотрудничество между компаниями представленной отрасли промышленности:

- возможность поиска партнеров по бизнесу и обмена опытом с представителями смежных отраслей;
- маркетинговый анализ рынка округа;
- установление долгосрочных взаимовыгодных контактов, заключение договоров, контрактов и т. д.
- профессиональные деловые мероприятия с участием специалистов и представителей органов власти;
- профессиональные мероприятия: круглые столы, семинары по основным вопросам реализации федеральных и окружных программ.

7 мая — День для широкой общественности: презентационный день для широкого спектра посетителей:

- презентации, мастер-классы для посетителей выставки;
- демонстрация современных технологий, оборудования и услуг участниками выставки;
- рекламные и PR-акции участников мероприятия;
- ознакомление с ценами товаров и услуг.

Приглашаем Вас принять участие в работе выставки и в деловой программе проекта. Рассмотрим Ваши предложения о сотрудничестве.

По вопросам участия обращаться:
КВЦ «Югра-Экспо», г. Ханты-Мансийск, ул. Студенческая, 19.
8(3467)35-95-86, expo_energy@mail.ru.

CONTENTS

Kudryavtsev K. Ya. <i>Spectral Method of Search of Keywords in Full-Text Databases</i>	2
---	---

The original approach of search of keywords in the full-text databases, grounded on the theory of spectral analysis and usage of algorithm of fast Fourier transform is offered.

Keywords: full-text databases, fast Fourier transform, keyword search in full-text databases

Polishuk Yu. V., Chernuh T. A. <i>The Construction of the Data Warehouse Oriented to the Operation with Quasistructured Informational Filling Documents</i>	9
--	---

The method to the construction of the data warehouse oriented to the operation with electronic documents, which informational filling is quasistructured, is considered. The mathematical interpretation of the method is proposed.

Keywords: the information storage, the quasistructured data, the automated information systems

Dryuchenko M. A., Sirota A. A. <i>Mathematical Algorithms and their Implementation for Staganalysis Information Detection Based on Statistical and Neural Network Algorithms</i>	14
---	----

A software complex for the steganographically-hidden data detection in files of different formats is developed. The possibility of integration of neural network algorithms with a well-known statistical methods of steganalys is examined.

Keywords: steganography, steganalys, statistical methods, change point, software complex

Kustov G. A., Nikolaeva M. A., Zotova O. F., Sharapov R. A. <i>Algorithms of Risk Decrease Preventive Measures Selection</i>	22
---	----

This article considers the approach to the risk analysis, risk value and risk management by an example of information risks. The logical-and-probabilistic method is the base of this approach. Information risks preliminary analysis using map of risks gave an opportunity to separate the most significant dangerous states of recourses the risk of which is not ignored and is not passed to control, but it is reduced due to the preventive measures — that are counter-measures. The authors offer their own algorithms of effective counter-measure set forming. Selection of each of them is defined by the initial information. These algorithms are the result of using the logical-and-probabilistic method for the purpose of risk identification and estimation.

Keywords: risk analysis and risk value, information risks, logical-and-probabilistic method, dangerous states, threat, threat significance, counter-measures, algorithm of effective counter-measure set forming

Zhernakov S. V. <i>Restoration of Loosed Information from Gas-Turbine Engine Sensors Under Conditions of On-Board Implementation with Use of Neural Network Technology</i>	28
---	----

One of approaches to restoration of loosed information from gas-turbine engine under conditions of on-board implementation on the base of neural network paradigm is considered. The technique of solving such tasks in the neural network basis is presented.

Keywords: fault-tolerance, checking, diagnosis, neural networks

Gluschenko O. I., Gorbatkov S. A. <i>Neural Model for Decision Support in the Municipal Budgeting</i>	34
--	----

In given article is designed economic concept of the efficient short-term planning the expense part of budget of the municipal formation and acceptance corresponding to management decisions, which allows to stimulate the objects lower level municipal institutions to increasing of efficiency of its work. The applying methods of the system analysis and neuro network of the toolboxes for modeling such complex problem as municipal budget.

Keywords: municipal budgeting, fund of material development, the neural network model, subventions, fiscal relations, municipalities, municipal institutions, budget service

Sobolev S. S., Sotov L. S., Harin V. N. <i>The Algorithm and Model of the Functional Generator of Permutations</i>	41
---	----

The algorithm and model of the high-efficiency generator of permutations of numbers (0, 1, ..., n) is offered.

The generator has matrix structure and hardware complexity $C_N = \frac{n}{2} (\log_2^2(n) - \log_2(n))$ of conditional logic elements that offers advantages at formation of shifts for great values n. The generator of permutations offered in work is intended for system of hardware formatting of binary lines of representation of the information in the COMPUTER.

Keywords: the generator of permutations, the generator of combinations, banyan switch, dynamic permutations, pseudorandom permutations

Abdullaev N. T., Aliev T. A., Dyshin O. A. Algorithm of Calculation of the Statistical Estimation of the Generalized Dimensions of Sets	46
--	----

On the basis of a method of a finding of all valid roots of the nonlinear equations the algorithm of construction of a statistical estimation of the generalized dimensions and a spectrum singularity, used for characteristics fractals and strange atrakt is offered.

Keywords: generalized dimensionality, set, fractals, estimation, multifractal analysis

Lvovich Ya. E., Korolev E. N. Advantages of the Object-Oriented Approach to Designing Systems of Training in Intellectual Systems of Decision-Making	56
---	----

The basic advantages of the object-oriented approach are described at designing complex systems of training. Some graphic designs of language of visual modelling of training systems are offered. These designs are applied to construction of conceptual, logic and graphic models of complex systems of training in intellectual systems of decision-making. They reflect both static, and dynamic aspects of functioning of training systems.

Keywords: designing of training systems, object-oriented technologies in training, visual modelling of training systems, Enterprise-components of training

Panteleev E. R., Karpov Ya. E. Development and Interpretation of Engineering Calculation Tasks Solution in a Web-Learning Environment	59
--	----

We consider methods and tools aimed at development and interpretation of the engineering calculation tasks solution in a Web-learning environment. Task representation model based on the Colored Petri Net formalism and network standards of its implementation that make proposed tools reusable in any contemporary operational and programming environment are justified. Example of development and interpretation of the engineering calculation task solution in the subject area of electro-mechanics is given.

Keywords: Web-learning, engineering calculation task, development and interpretation tools

Safronov V. V. The Branches and Borders Method in Direct and Inverse Problems of the Disciplines Cycle Study Time Optimal Assignment	62
---	----

The direct and inverse problem settings of the disciplines study time valid selection on higher vocational training of specialists are handled in this article. The branches and borders method is offered for such a problems decision. The features of estimations construction at branching levels are showed. The numerical examples are given.

Keywords: the state educational institution, the branches and borders method, the estimations, the disciplines cycle study time

Gurevich I. M. Information as a Universal Inhomogeneity	66
--	----

There is still no universally accepted definition of information, although known to many dozens of definitions. This is due, above all, the number and variety of subject areas in which the notion of information. The notion of information is used in almost all fields of science, culture and economy. In this paper, an another attempt to provide a general definition of information. More specifically, the author formalizes and justifies the universality of a well-known definitions. You can see examples from different subject areas.

Keywords: homogeneity, heterogeneity, information, information entropy, information divergence, the volume of information

Адрес редакции:

107076, Москва, Стромынский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *О. А. Ефремова*.

Корректор *Т. В. Пчёлкина*

Сдано в набор 05.02.2010. Подписано в печать 16.03.2010. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.

Усл. печ. л. 9,8. Уч.-изд. л. 10,67. Заказ 294. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Отпечатано в ООО "Подольская Периодика"

142110, Московская обл., г. Подольск, ул. Кирова, 15
